心理学报 2023, Vol. 55, No. 9, 1573−1586
Acta Psychologica Sinica

https://doi.org/10.3724/SP.J.1041.2023.01573

联合作答精度和作答时间的概率态认知诊断模型*

田亚淑 詹沛达 王立君

(浙江师范大学心理学院；浙江省儿童青少年心理健康与心理危机干预智能实验室；
浙江省智能教育技术与应用重点实验室，金华 321004)

摘 要 对多模态数据的联合分析是改进结果评价、健全综合评价的主要途径。针对概率态认知诊断模型(CDM)仅能分析题目作答精度(RA)的局限，本文基于联合−层级建模框架和联合−交叉负载建模框架提出三个可联合分析 RA 和题目作答时间(RT)的概率态联合 CDM。模拟研究和实证研究结果表明：(1)新模型参数估计返真性良好，额外引入 RT 有助于提高参数估计精度并提供有关个体加工速度的测量；(2)基于联合−交叉负载建模框架构建的模型对测验情境的兼容性优于基于联合−层级建模框架构建的模型；(3)概率态属性比确定态属性更精细地反映个体对属性的掌握情况。

关键词 认知诊断，概率态属性，题目作答时间，联合建模框架，交叉负载

分类号 B841

1 引言

认知诊断测评可以提供有关学生知识结构或加工技能(统称为“(潜在)属性”)的诊断信息；可为教师实施补救教学或有针对性干预提供参考，有助于促进学生发展(Tang & Zhan, 2021)。认知诊断模型(cognitive diagnosis model, CDM)是刻画属性与外显行为之间关系的统计模型，其建构的合理性(如，模型与测试情境的匹配度)影响诊断结果的准确性和可解释性。目前，大致可将属性分为两种：确定态属性(deterministic attribute)和概率态属性(probabilistic attribute)：前者将被试的属性掌握状态诊断为确定的类别(如，“1”表示掌握，“0”表示未掌握)；而后者将被试对属性的掌握诊断为 0 到 1 的连续状态，用于量化被试对属性的掌握概率(如，“0.8”表示掌握概率为 80%) (Zhan, Wang, et al., 2018)。Zhan (2021)进一步指出概率态属性是“个体掌握某种属性的说法是正确的概率(the probability that the statement that a person masters an attribute is true)”，即概率态属性并没有否定属性的二分特性，只是从概率视角对其进行解读和建模。相比于确定态属性，概率态属性能更精细地区分被试间的个体差异(詹沛达，田亚淑 等，2020)；尤其是在描述被试发展情况时，基于概率态属性的反馈比基于确定态属性的反馈更精细地反映学生的发展变化(Zhan, 2021)，更有益于肯定学生的努力。

针对不同的测验情境和理论假设，研究者们提出了不同的概率态 CDM (毛秀珍，2014；詹沛达，边玉芳，2015；Liu et al., 2018；Zhan, Wang, et al., 2018; Zhan, 2021)。然而，现有的概率态 CDM 仅能分析单一模态数据——题目作答精度(response accuracy, RA)，忽略了诸如题目作答时间(response times, RTs)、鼠标点击次数和行动序列等过程数据。其中，RT 作为一种有代表性的过程数据，是指被试作答每道题目花费的时间；可以反映被试的(潜在)加工速度。目前，无论是大规模测评项目[如，国际学生

收稿日期：2022-08-30
* 国家自然科学基金青年基金项目(31900795)资助。
通信作者：詹沛达，E-mail: pdzhan@gmail.com；王立君，E-mail: frankwlj@163.com

评估项目(PISA)、美国国家教育进展评估(NAEP)和国际数学与科学趋势研究(TIMSS)], 还是一些具有实验性质的小规模测评(如, 游戏化测评), 记录RT已经成为一种常态。

近些年, 研究者们开发了一系列RT分析模型(郭磊 等, 2017; de Boeck & Jeon, 2019)。已有研究表明数据分析时额外引入RT, 有助于提高被试参数估计的精度(Bolsinova & Tijmstra, 2018; Zhan, Jiao, & Liao, 2018), 并有助于识别被试的异常作答行为(Wang & Xu, 2015); 通过RT所反映的加工速度还可以进一步探索被试的认知风格(如, Yan, 2010), 丰富诊断反馈所包含的信息(Zhan et al., 2022)。鉴于在数据分析中引入RT的诸多优势, 如何将RT引入概率态CDM以进一步提高参数估计精度并丰富诊断反馈信息, 是一个有待解决的方法学问题。

目前, 在认知诊断测评中, 联合−层级认知诊断建模框架(Zhan, Jiao, & Liao, 2018)是同时分析RA和RT的主要框架之一, 如图1所示。在该框架中, RA用于测量被试的属性和(潜在)能力、RT用于测量被试的加工速度。该框架具有较高灵活性, 可通过替换该框架中的测量模型实现不同的数据分析需求(Huang, 2020; Peng et al., 2022; Zhan et al., 2022; 詹沛达, 2022; 郑天鹏 等, 2023)。目前, 该框架下的所有模型都是针对确定态属性构建的, 难以精细化区分被试间的个体差异或提供精细化反馈。鉴于该框架的灵活性, 可尝试将概率态CDM引入该框架, 构建可联合分析RA和RT的联合−层级概率态CDM, 这是本研究拟解决的第一个主要问题。

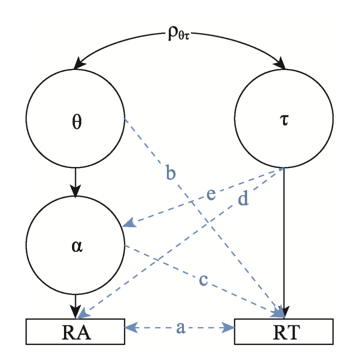


图 1 联合−层级认知诊断建模框架中条件独立性假设示意图

注：RA = 作答精度; RT = 作答时间; θ = 能力; τ = 加工速度; α = 属性; ρ = 能力与加工速度的相关系数; 虚线表示模型的条件独立假设: a = 给定能力和加工速度后, RT和RA条件独立; b = 给定加工速度后, 能力和RT条件独立; c = 给定能力和加工速度后, 属性和RT条件独立; d = 给定能力后, 加工速度和RA条件独立; e = 给定能力后, 加工速度和属性条件独立.

联合−层级认知诊断建模框架作为联合−层级建模框架(van der Linden, 2007)在认知诊断测评中的应用, 尽管得到了较广泛的研究支持, 但基于该框架所构建的联合模型至少需要满足5个条件独立性假设才能够合理、准确地对数据进行分析和解释。如图1所示：

(1)给定能力和加工速度后, RT和RA之间条件独立(图1中a);

(2)给定加工速度后, 能力和RT之间条件独立(图1中b);

(3)给定能力后和加工速度后, 属性和RT之间条件独立(图1中c);

(4)给定能力后, 加工速度和RA之间条件独立(图1中d);

(5)给定能力后, 加工速度和属性之间条件独立(图1中e)。

然而, 有研究发现实际测验中存在违背上述假设的情况。比如, Meng等人(2015)发现RT和RA之间存在一个变量(如, 题目难度)使得两者有相依性; Bolsinova等人(2017)尝试对RA和RT之间的条件独立性进行建模; Bolsinova和Tijmstra (2018)释放能力和RT之间条件独立性假设后发现可以进一步提高能力的估计精度。同时, 上述条件独立性假设也使得联合−层级认知诊断建模框架只能通过能力和加工速度之间的相关性(即, 图1中ρ)来获取RT中所包含的辅助信息, 进而达到提高诊断分类准确性的目的(詹沛达, 2022); 类似的理论缺陷在联合−层级建模框架中也存在(Ranger, 2013; Bolsinova & Tijmstra, 2018)。而在实际测验中, 可能存在能力和加工速度之间的相关系数较低, 甚至趋近于零的情况(Bolsinova et al., 2017); 此时基于联合−层级认知诊断建模框架所构建的模型将难以从RT中获取能力或属性的相关信息。针对上述局限, 詹沛达(2022)提出了联合−交叉负载(载荷)认知诊断建模框架, 该建模框架通过交叉负载直接利用RT为能力或属性提供信息, 释放了联合−层级认知诊断建模框架的部分条件独立性假设(图1中b或c), 增加了建模框架的适用范围。郑天鹏等人(2023)在联合−层级认知诊断建模框架的基础上尝试释放了图1中条件独立性d, 认为被试的加工速度会影响其正确作答概率。虽然释放图1中任一条件独立性假设路径在理论上都是可以的, 但考虑到联合建模的主要目的之一是为了促进对核心特质(即能力)的估计精度, 所以释放b或c路径是一种更常见的选择

(Bolsinova & Tijmstra, 2018; 詹沛达, 2022)。基于此, 本研究拟在所构建的联合–层级概率态 CDM 的基础上, 借鉴联合–交叉负载认知诊断建模框架, 进一步构建联合–交叉负载概率态 CDM (joint-cross-loading CDM for probabilistic attributes), 这是本研究拟解决的第二个主要问题。

针对上述两个研究问题, 本研究旨在丰富概率态 CDM 的可分析数据类型, 并为以精细化诊断作为数据分析目标的测评提供框架支持。在实践中, 全面且精准地了解学生的学习现状是因材施教的必要前提; 使用联合概率态 CDM 不仅可以实现对学生学习状态的精细化区分, 为因材施教和精准教学落地提供数据支持, 还可以提供有关学生加工速度的分析结果, 有助于了解学生的认知或学习风格。

下文首先简单回顾联合–层级和联合–交叉负载认知诊断建模框架, 以及一个有代表性的概率态 CDM——高阶概率态输入, 噪音连接(higher-order probabilistic-inputs, noisy conjunctive, HO-PINC)模型(Zhan, Wang, et al., 2018)。其次, 依次通过两个模拟研究分别阐述并探究基于上述两框架所构建的三个新模型。然后, 通过第三个模拟研究来交叉对比三个新模型的表现。再然后, 以一则 PISA 2012 计算机化数学测验数据为例来展现三个新模型的实践可应用性及相对优势。最后, 总结研究并指出研究局限及未来的拓展方向。

2　两个联合认知诊断建模框架与概率态认知诊断模型

图 2 呈现了两种联合认知诊断建模框架的示意图, 其中图 2(a)为联合–层级认知诊断建模框架(Zhan, Jiao, & Liao, 2018)。该建模框架包含两层模型: 第一层为测量模型, 比如以高阶 DINA(de la Torre & Douglas, 2004)作为 RA 的测量模型, 并以对数正态 RT 模型(lognormal RT model, LRTM) (van der Linden, 2006)作为 RT 的测量模型; 第二层为结构模型, 采用二元正态分布描述能力与加工速度之间的关系。图 2(b)和图 2(c)为联合–交叉负载认知诊断建模框架下的两种建模方式(詹沛达, 2022): 前者基于能力, 假设 RT 先对能力提供辅助信息(即被试的能力高低会影响 RT), 进而间接影响属性的分类准确性; 后者基于属性, 假设 RT 直接对属性提供辅助信息(即被试的属性掌握情况会影响 RT)。

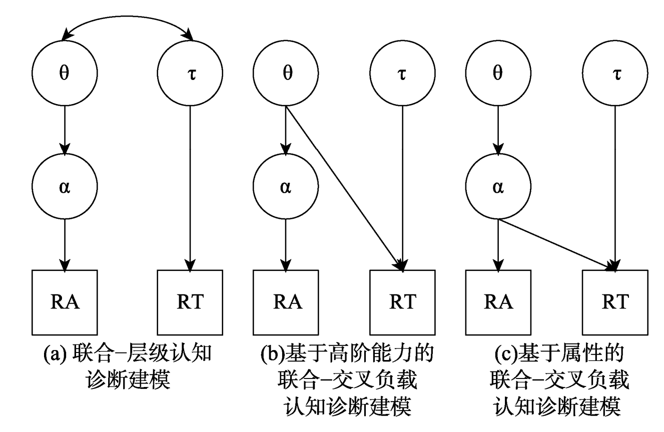


图 2　联合–层级和联合–交叉负载认知诊断建模示意图

注: RA = 作答精度; RT = 作答时间; θ = 能力; τ = 加工速度; α = 属性.

目前两建模框架中属性均为确定态属性(即二分属性), 进而能力与属性之间的关系被定义为(de la Torre & Douglas, 2004):

$$\text{logit}(Prob(\alpha_{nk}=1\mid\theta_n,\lambda_{0k},\lambda_{1k}))=\lambda_{1k}\theta_n+\lambda_{0k}\,,\quad(1)$$

其中, α_{nk} 为被试 n 对属性 k 的掌握类别, $\alpha_{nk}\in\{0,1\}$; θ_n 为被试 n 的能力; λ_{0k} 和 λ_{1k} 分别为属性 k 的截距和区分度参数。

与确定态属性不同, 概率态属性常被赋值为一个 0 到 1 之间的(概率)数值。考虑到属性之间的相关性, Zhan, Wang 等(2018)将概率态属性构建为:

$$\text{logit}(m_{nk})=\lambda_{1k}\theta_n+\lambda_{0k}\,,\quad(2)$$

其中, 概率态属性 m_{nk} 可反映被试 n 对属性 k 的掌握概率, $m_{nk}\in(0,1)$; 其他参数含义同上。

Zhan (2022)进一步指出 m_{nk} 和 $Prob(\alpha_{nk}=1\mid\theta_n)$ 尽管有相同的表达式, 但由于嵌入测量模型的方式不同, 导致两者在解释上具有差异性。前者可被解释为被试对属性的掌握概率, 其数值变化会直接影响被试的正确作答概率。而后者可被解释为被试归入“掌握”类的确定性程度, 类似于连续变量估计中的标准误, 反映的是分类的误差或确定性程度(Bradshaw & Levy, 2019); 因其数值变化并不影响被试的正确作答概率, 仅影响分类的确定性, 所以不应被解释为被试对属性的掌握概率(Zhan, 2021)。

基于公式 2, HO-PINC 的题目作答函数(item response function, IRF)可以被构建为:

$$Prob(X_{ni}=1\mid m_n,g_i,s_i,q_{ik})=g_i+(1-s_i-g_i)\prod_{k=1}^{K}m_{nk}^{q_{ik}}\,,\quad(3)$$

式中, g_i 和 s_i 分别为题目 i 的猜测参数和失误参数; q_{ik} 为 Q 矩阵(Tatsuoka, 1983)中元素, $q_{ik}=1$ 表示题目 i 考查属性 k, 反之 $q_{ik}=0$; $\prod_{k=1}^{K}m_{nk}^{q_{ik}}$ 为被试 n 对

题目 i 的理想作答概率，表示同时掌握题目 i 所需的多个属性的联合概率；其他参数含义同上。

3 研究 1：针对作答精度和时间的联合−层级概率态认知诊断模型

3.1 模型构建

遵循联合−层级认知诊断建模框架，新模型共包含两层模型。第一层为测量模型，其中 RA 模型采用 HO-PINC (公式 3), RT 模型采用 LRTM：

$$logT_{ni} \sim N(\xi_i - \tau_n, \sigma_i^2), \quad (4)$$

式中，T_{ni} 为被试 n 作答题目 i 所花费的时间；τ_n 为被试 n 的加工速度；ξ_i 和 $1/\sigma_i$ 分别为题目 i 的时间强度参数和时间精度参数。第二层为结构模型，包含用于描述能力和加工速度之间关系的和用于描述题目参数之间相依性的两个多元正态分布，分别为：

$$\begin{pmatrix}\theta_n\\ \tau_n\end{pmatrix} \sim MVN\left(\begin{pmatrix}0\\0\end{pmatrix}, \boldsymbol{\Sigma}_{person}\right), \boldsymbol{\Sigma}_{person} = \begin{pmatrix}\sigma_\theta^2 & \\ \sigma_{\theta\tau} & \sigma_\tau^2\end{pmatrix}, \quad (5)$$

和

$$\begin{pmatrix} logit(g_i) \\ logit(1-s_i)-logti(g_i) \\ \xi_i \end{pmatrix} \equiv \begin{pmatrix} g_i^{\sim} \\ s_i^{\sim} \\ \xi_i \end{pmatrix} \sim MVN\left(\begin{pmatrix}\mu_{g^{\sim}}\\ \mu_{s^{\sim}}\\ \mu_\xi\end{pmatrix}, \boldsymbol{\Sigma}_{item}\right),$$

$$\boldsymbol{\Sigma}_{item} = \begin{pmatrix}\sigma_{g^{\sim}}^2 & & \\ \sigma_{s^{\sim}g^{\sim}} & \sigma_{s^{\sim}}^2 & \\ \sigma_{\xi g^{\sim}} & \sigma_{\xi s^{\sim}} & \sigma_\xi^2\end{pmatrix}, \quad (6)$$

式中，$g_i^{\sim}$ 和 $s_i^{\sim}$ 分别为转换到 logit 量尺上的猜测参数和失误参数，以保证可以与 ξ_i 构建多元正态分布(Zhan, Jiao, & Liao, 2018; Zhan et al., 2022)。综上，本研究将公式 2~6 称为针对作答精度和时间的联合−层级 PINC (JRT-PINC)模型。

基于联合−层级建模框架，为使模型可识别，JRT-PINC 需要满足如下条件独立性假设：

(1) 给定 θ_n 后，各 m_{nk} 满足条件独立；

(2) 给定 m_{nk} 后，各 X_{ni} 满足条件独立；

(3) 给定 τ_n 后，各 $\log T_{ni}$ 满足条件独立；

(4) 给定 θ_n 和 τ_n 后，X_{ni} 和 $\log T_{ni}$ 满足条件独立；

(5) 给定 θ_n 后，各 τ_n 和 m_{nk} 满足条件独立；

(6) 给定 θ_n 和 m_{nk}，各 τ_n 和 X_{ni} 满足条件独立；

(7) 给定 θ_n 和 τ_n 后，m_{nk} 和 $\log T_{ni}$ 满足条件独立；

(8) 给定 τ_n 后，θ_n 和 $\log T_{ni}$ 满足条件独立。

其中，前 3 个条件独立性假设是测量模型自身所需的，后 5 个条件独立性假设是联合−层级建模框架所需的(同引言)。

3.2 参数估计

本研究采用全贝叶斯 MCMC 算法对 JRT-PINC 进行参数估计，使用 R 软件中的 R2jags 包(Version 0.7-1)调取 JAGS 软件(Version 4.3.0)实现参数估计。网络版附录 S1 节对比了新模型在有经验信息先验下和低信息先验下的表现，两种情况下结果一致性较高，表明新模型对包含不同信息量的先验分布具有一定的鲁棒性。新模型的参数估计 JAGS 代码及示例数据见 https://osf.io/hys7c/?view_only=cb357a6f5032424ab36b7fbda6df4d40，关于如何使用 JAGS 进行贝叶斯参数估计可参见 Zhan 等(2019)。

3.3 模拟研究

3.3.1 数据生成

模拟研究包括 4 个自变量：(1)样本量(N)：200 和 500; (2)题目数量(I)：15 和 30; (3)能力与加工速度的相关系数($\rho_{\theta\tau}$)：−0.5, −0.3, 0, 0.3 和 0.5; (4)数据分析模型：JRT-PINC 和 HO-PINC，用于探究额外引入 RT 对参数估计精度带来的影响。此外属性数(K)固定为 5 个，对应的 Q 矩阵见图 3。该 Q 矩阵包含两个可达矩阵，以保证其完备性和可识别性(Xu & Zhang, 2016)。研究选择的题目数量、样本量和其他相关参数均参考已有研究设置(如，詹沛达, 2022)。

题目参数依据公式 6 生成：所有题目猜测参数均值 $\mu_{g^{\sim}}$ = −2.197、所有题目失误参数均值 $\mu_{s^{\sim}}$ = 4.394 和所有题目时间强度参数均值 $\mu_\xi = 4$；各参数的方差协方差矩阵为：

$$\boldsymbol{\Sigma}_{item} = \begin{pmatrix}\sigma_{g^{\sim}}^2 & & \\ \sigma_{s^{\sim}g^{\sim}} & \sigma_{s^{\sim}}^2 & \\ \sigma_{\xi g^{\sim}} & \sigma_{\xi s^{\sim}} & \sigma_\xi^2\end{pmatrix} = \begin{pmatrix}1 & & \\ -0.8 & 1 & \\ -0.25 & 0.15 & 0.25\end{pmatrix},$$

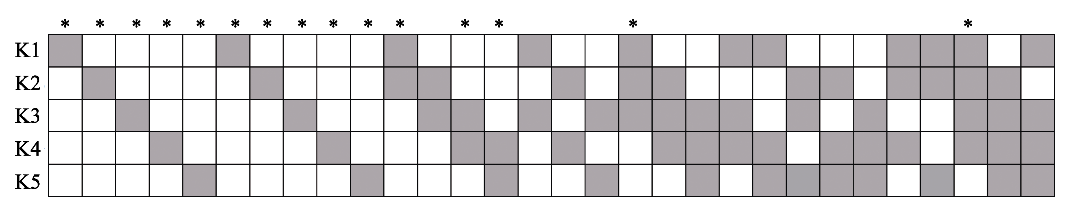


图 3 研究 1 中模拟研究的 Q 矩阵

注：灰色为“1”、白色为“0”；标记*的题目为 I = 15

表 1 研究 1 中能力和加工速度参数估计返真性

N	I	$\rho_{\theta\tau}$	JRT-PINC						HO-PINC		
			θ			τ			θ		
			Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor
200	15	−0.5	−0.001	0.458	0.887	0.000	0.121	0.896	−0.002	0.479	0.875
		−0.3	−0.003	0.468	0.881	0.001	0.122	0.896	−0.002	0.480	0.875
		0.0	−0.003	0.473	0.879	−0.000	0.122	0.949	−0.004	0.480	0.875
		0.3	−0.001	0.469	0.881	−0.001	0.121	0.899	−0.003	0.480	0.874
		0.5	−0.000	0.458	0.887	−0.000	0.120	0.902	−0.001	0.483	0.874
	30	−0.5	−0.001	0.397	0.915	−0.000	0.085	0.953	−0.004	0.412	0.908
		−0.3	0.000	0.403	0.913	−0.001	0.086	0.953	−0.004	0.411	0.909
		0.0	−0.002	0.405	0.911	−0.000	0.086	0.975	−0.004	0.413	0.908
		0.3	−0.002	0.397	0.915	−0.002	0.086	0.954	−0.004	0.410	0.909
		0.5	−0.003	0.389	0.919	−0.001	0.085	0.953	−0.005	0.411	0.909
500	15	−0.5	−0.000	0.456	0.888	−0.000	0.121	0.902	0.000	0.476	0.876
		−0.3	0.000	0.467	0.881	−0.001	0.122	0.901	0.000	0.477	0.876
		0.0	−0.000	0.473	0.878	−0.000	0.122	0.948	−0.001	0.476	0.875
		0.3	−0.001	0.470	0.880	−0.001	0.122	0.901	−0.001	0.478	0.875
		0.5	−0.000	0.460	0.885	−0.001	0.121	0.903	−0.001	0.478	0.876
	30	−0.5	0.000	0.391	0.918	−0.000	0.088	0.948	0.000	0.407	0.911
		−0.3	0.002	0.400	0.914	−0.001	0.088	0.948	0.000	0.407	0.911
		0.0	−0.001	0.402	0.913	−0.000	0.089	0.973	0.000	0.405	0.912
		0.3	0.000	0.397	0.915	−0.001	0.088	0.948	0.000	0.405	0.912
		0.5	0.001	0.387	0.920	0.000	0.088	0.948	0.000	0.405	0.912

注: JRT-PINC = 联合−层级概率态输入, 噪音连接模型; HO-PINC = 高阶概率态输入, 噪音连接模型; θ = 能力; τ = 加工速度; N = 样本量; I = 题目数量; $\rho_{\theta\tau}$ = 能力与加工速度的相关系数; Bias = 所有被试的平均偏差; RMSE = 所有被试的平均均方根误差; Cor = 估计值与真值之间的相关系数.

另外, 设定所有题目的时间精度参数均为 $1/\sigma_i = 2$。

被试参数依据公式 5 生成, 其中, 能力的方差 $\sigma_\theta^2 = 1$ 且加工速度的方差 $\sigma_\tau^2 = 0.15$。被试的概率态属性依据公式 2 生成, 设定所有属性的区分度参数为 $\lambda_{1k} = 1.5$, 属性截距参数为 $\boldsymbol{\lambda}_0 = (-1.0, -0.5, 0.0, 0.5, 1.0)'$, 此时各属性间为中等相关。最终, 根据 JRT-PINC 生成所有被试在所有题目上的 RA 和 RT。为减少随机误差, 每种模拟条件下各生成 30 组数据。

3.3.2 分析

分别使用 JRT-PINC 和 HO-PINC 分析该数据。分析采用两条马尔可夫链(初始值随机), 每条链包含 10,000 次迭代, 前 5,000 次用于预热(burn-in)。采用潜在量尺缩减因子(potential scale reduction factor, PSRF)检验各参数是否收敛, PSRF < 1.1 或 1.2 表示参数已收敛(de la Torre & Douglas, 2004; Zhan, Jiao, & Liao, 2018)。使用后验均值作为贝叶斯参数估计的“点”估计结果。使用偏差(bias)、均方根误差(root mean square error, RMSE)和皮尔逊积差相关系数(Cor)作为参数估计返真性指标。

3.3.3 结果

结果显示所有参数的 PSRF 均小于 1.2, 表示各参数均已收敛。表 1 呈现了能力参数和加工速度参数的返真性。首先, JRT-PINC 中能力参数(θ)的估计返真性优于 HO-PINC 的, 表明额外引入 RT 所包含信息有助于提高能力参数的估计精度。其次, JRT-PINC 的能力参数和加工速度参数(τ)在所有模拟条件下的返真性均表现较好, 且 Cor 指标反映出加工速度参数的返真性优于能力参数的。[1]再有, 对 JRT-PINC 而言, 能力与加工速度的相关系数($\rho_{\theta\tau}$)越高, 能力参数的返真性越好, 但加工速度参数的返真性不受影响; 该结果与已有联合模型研究的发现一致, 即两特质之间的相关系数越高, 越有助于能力参数汲取 RT 中包含的信息。且即便是在 $\rho_{\theta\tau} = 0$

[1]由于能力参数和加工速度参数的真值的方差不同(即量尺不同), 难以直接通过 Bias 和 RMSE 比较两者的返真性优劣; 而 Cor 指标是从参数估计值和真值的排序一致性角度反映参数估计返真性的, 并不受量尺差异的影响。

的条件下, JRT-PINC 中能力参数的返真性仍略优于 HO-PINC 中的; 这可能由于参数估计时 JRT-PINC 仍假设能力会通过潜在结构(二元正态分布)去利用 RT 所包含的信息, 进而自由估计两者之间的协方差导致的。此外, $\rho_{\theta\tau}$ 的正负号对结果没有影响。整体而言, 题目数量越多, 能力与加工速度相关系数越高, JRT-PINC 的被试参数返真性越好; 样本量提高也有助于提高被试参数的返真性, 但影响较小; 这意味着 200 被试的样本量足以满足 JRT-PINC 被试参数估计的要求。

表 2 呈现了属性参数的返真性(bias 见网络版附录中表 S2.1)。首先, JRT-PINC 在各条件下的返真性均较好。题目数量越多, 能力与加工速度相关系数越高, 则属性的返真性越好; 同样, 样本量提高也有助于提高属性的返真性, 但影响较小。其次, 各条件下 JRT-PINC 的返真性均优于 HO-PINC 的; 同样, 即便是在 $\rho_{\theta\tau} = 0$ 的条件下, JRT-PINC 中属性的返真性也略优于 HO-PINC 中属性的返真性。另外, 不同属性的返真性与属性截距参数有关, 但影响不大, 基本趋势是: 属性截距参数越低(即掌握属性难度越大), 则返真性越好。

题目参数和方差协方差矩阵参数估计返真性见网络版附录 S2。整体而言, JRT-PINC 的题目参数返真性优于 HO-PINC 的; JRT-PINC 中各题目参数在各模拟条件下的返真性均表现较好, 且具有较一致的变化趋势: 样本量和题目数量越大, 题目参数的返真性越好, 而能力与加工速度的相关系数影响较小。

4　研究 2: 针对作答精度和时间的联合−交叉负载概率态认知诊断模型

基于图 2(b)和 2(c), 研究 2 通过两个子研究分别构建基于能力的联合−交叉负载 PINC (CJRT-PINC-θ)模型(子研究 1)和基于属性的联合−交叉负载 PINC (CJRT-PINC-*m*)模型(子研究 2)。

表 2　研究 1 中概率态属性参数估计的返真性

N	*I*	$\rho_{\theta\tau}$	RMSE										Cor									
			JRT-PINC					HO-PINC					JRT-PINC					HO-PINC				
			m_1	m_2	m_3	m_4	m_5	m_1	m_2	m_3	m_4	m_5	m_1	m_2	m_3	m_4	m_5	m_1	m_2	m_3	m_4	m_5
200	15	−0.5	0.129	0.137	0.139	0.145	0.139	0.155	0.152	0.152	0.155	0.158	0.899	0.901	0.896	0.888	0.875	0.888	0.890	0.886	0.877	0.861
		−0.3	0.130	0.136	0.144	0.150	0.141	0.158	0.148	0.153	0.158	0.160	0.895	0.897	0.892	0.883	0.869	0.887	0.891	0.885	0.876	0.860
		0	0.132	0.137	0.144	0.148	0.141	0.161	0.147	0.155	0.157	0.157	0.894	0.896	0.890	0.881	0.867	0.886	0.891	0.885	0.876	0.861
		0.3	0.127	0.138	0.145	0.151	0.143	0.154	0.148	0.159	0.158	0.157	0.896	0.896	0.890	0.882	0.869	0.887	0.891	0.884	0.876	0.862
		0.5	0.128	0.139	0.144	0.147	0.142	0.159	0.151	0.158	0.157	0.161	0.898	0.899	0.894	0.886	0.874	0.884	0.889	0.883	0.874	0.860
	30	−0.5	0.097	0.108	0.124	0.127	0.133	0.117	0.125	0.142	0.144	0.151	0.946	0.940	0.928	0.912	0.890	0.940	0.933	0.920	0.904	0.881
		−0.3	0.100	0.110	0.126	0.128	0.132	0.120	0.124	0.142	0.143	0.148	0.945	0.939	0.927	0.911	0.887	0.940	0.933	0.921	0.905	0.882
		0	0.103	0.110	0.128	0.130	0.134	0.122	0.126	0.142	0.148	0.153	0.944	0.938	0.925	0.909	0.885	0.939	0.933	0.920	0.903	0.879
		0.3	0.097	0.110	0.123	0.131	0.129	0.121	0.126	0.140	0.148	0.150	0.946	0.939	0.928	0.912	0.890	0.940	0.933	0.921	0.905	0.881
		0.5	0.096	0.110	0.124	0.127	0.126	0.120	0.127	0.143	0.144	0.148	0.947	0.940	0.929	0.915	0.894	0.940	0.933	0.921	0.905	0.881
500	15	−0.5	0.111	0.131	0.130	0.139	0.139	0.132	0.140	0.141	0.145	0.160	0.906	0.905	0.901	0.890	0.874	0.896	0.896	0.892	0.879	0.861
		−0.3	0.113	0.132	0.134	0.140	0.142	0.132	0.140	0.141	0.143	0.163	0.902	0.901	0.896	0.885	0.867	0.896	0.895	0.892	0.879	0.861
		0	0.113	0.135	0.135	0.142	0.144	0.133	0.140	0.140	0.145	0.165	0.900	0.898	0.894	0.883	0.864	0.895	0.895	0.892	0.880	0.861
		0.3	0.114	0.136	0.134	0.142	0.146	0.132	0.143	0.140	0.145	0.165	0.900	0.899	0.895	0.884	0.866	0.894	0.894	0.891	0.879	0.860
		0.5	0.111	0.134	0.132	0.139	0.140	0.131	0.140	0.141	0.144	0.163	0.904	0.902	0.900	0.889	0.872	0.894	0.895	0.891	0.879	0.860
	30	−0.5	0.088	0.098	0.107	0.119	0.125	0.101	0.107	0.121	0.126	0.145	0.949	0.943	0.932	0.916	0.895	0.945	0.938	0.926	0.908	0.884
		−0.3	0.087	0.101	0.109	0.121	0.128	0.100	0.109	0.122	0.126	0.145	0.947	0.941	0.929	0.913	0.890	0.944	0.938	0.925	0.908	0.884
		0	0.090	0.100	0.109	0.121	0.128	0.101	0.108	0.121	0.125	0.144	0.947	0.941	0.929	0.912	0.889	0.945	0.939	0.926	0.910	0.886
		0.3	0.089	0.098	0.109	0.120	0.126	0.100	0.105	0.123	0.125	0.145	0.948	0.942	0.930	0.915	0.892	0.945	0.939	0.926	0.910	0.886
		0.5	0.089	0.095	0.107	0.116	0.123	0.100	0.105	0.122	0.125	0.141	0.949	0.944	0.933	0.919	0.898	0.945	0.939	0.927	0.910	0.886

注: JRT-PINC = 联合−层级概率态输入, 噪音连接模型; HO-PINC = 高阶概率态输入, 噪音连接模型; θ = 能力; τ = 加工速度; N = 样本量; I = 题目数量; ρ = 能力与加工速度的相关系数。

4.1 子研究 1：CJRT-PINC-θ

4.1.1 模型构建

基于图 2(b)，CJRT-PINC-θ 假设被试的能力变化会影响其作答该题目的耗时，即释放了 JRT-PINC 中能力和 RT 之间的条件独立性假设(图 1 中 b)。CJRT-PINC-θ 同样包含两层模型，其中 RA 的测量模型为 HO-PINC 模型(公式 3)，RT 的测量模型为：

$$logT_{ni} \sim N(\xi_i - \tau_n + \varphi_i\theta_n, \sigma_i^2), \quad (7)$$

式中，φ_i 为交叉负载，表示 θ_n 对 $\log T_{ni}$ 的影响大小：当 $\varphi_i > 0$ 时，θ_n 越大则 RT 越长，反之当 $\varphi_i < 0$ 时，θ_n 越大则 RT 越短。另外，为了保证模型可识别性(即 θ_n 和 τ_n 的相关性已经被交叉负载解释)，该模型假设 θ_n 和 τ_n 为独立分布：

$$\theta_n \sim N(0, \sigma_\theta^2), \tau_n \sim N(0, \sigma_\tau^2). \quad (8)$$

另外，题目参数之间关系被描述为：

$$\begin{pmatrix} g_i^{\sim} \\ s_i^{\sim} \\ \xi_i \\ \varphi_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{g^{\sim}} \\ \mu_{s^{\sim}} \\ \mu_\xi \\ \mu_\varphi \end{pmatrix}, \boldsymbol{\Sigma}_{item}\right), \boldsymbol{\Sigma}_{item} = \begin{pmatrix} \sigma_{g^{\sim}}^2 & \cdots & \\ \vdots & \ddots & \vdots \\ \sigma_{\varphi g^{\sim}} & \cdots & \sigma_\varphi^2 \end{pmatrix}, \quad (9)$$

式中各参数含义同上。仍采用贝叶斯 MCMC 算法对 CJRT-PINC-θ 进行参数估计，各待估计参数的先验分布见网络版附录 S1。

4.1.2 模拟研究：数据生成与结果

为了更符合实际测试的复杂情境，模拟研究设置不同题目中 θ_n 对 RT 的影响不同(詹沛达, 2022)：设定自变量交叉负载(φ_i)满足方差为 0.15 的正态分布，其中低影响效应 $\mu_\varphi = 0.1$，高影响效应 $\mu_\varphi = 0.5$。固定样本量 $N = 200$，题目数量 $I = 15$，其余参数设定与研究 1 相同。根据 CJRT-PINC-θ (公式 2~3 和公式 7~9)生成所有被试在所有题目上的 RA 和 RT。

使用 CJRT-PINC-θ 和 HO-PINC 分析该数据。结果显示所有参数的 PSRF 均小于 1.2，表示各参数均已收敛。表 3 呈现了能力参数和加工速度参数的估计返真性。首先，CJRT-PINC-θ 的返真性均优于 HO-PINC 的，表明额外引入 RT 所包含信息有助于提高参数估计精度。其次，随着交叉负载均值提高，能力参数的返真性有所提高但加工速度参数的返真性有所下降，这与詹沛达(2022)的发现一致。表 4 呈现了

表 3　研究 2 (子研究 1)中被试参数估计返真性

分析模型	μ_φ	θ Bias	θ RMSE	θ Cor	τ Bias	τ RMSE	τ Cor
CJRT-PINC-θ	0.1	−0.006	0.295	0.958	0.000	0.138	0.969
	0.5	−0.007	0.287	0.958	−0.001	0.194	0.971
HO-PINC	0.1	−0.007	0.481	0.871	–	–	–
	0.5	−0.008	0.503	0.864	–	–	–

注: CJRT-PINC-θ = 基于能力的联合−交叉负载概率态输入, 噪音连接模型; HO-PINC = 高阶概率态输入, 噪音连接模型; θ = 能力; τ = 加工速度; μ_φ = 交叉负载均值; Bias = 所有被试的平均偏差; RMSE = 所有被试的平均均方根误差; Cor = 估计值与真值之间的相关系数.

表 4　研究 2 (子研究 1)概率态属性参数估计返真性

分析模型	μ_φ	指标	m_1	m_2	m_3	m_4	m_5
CJRT-PINC-θ	0.1	Bias	−0.005	−0.005	−0.001	−0.014	−0.012
		RMSE	0.097	0.113	0.101	0.119	0.102
		Cor	0.953	0.952	0.953	0.947	0.948
	0.5	Bias	−0.004	−0.030	−0.006	−0.009	−0.002
		RMSE	0.097	0.118	0.107	0.111	0.098
		Cor	0.952	0.949	0.952	0.950	0.949
HO-PINC	0.1	Bias	0.041	0.004	−0.005	−0.037	−0.052
		RMSE	0.153	0.148	0.155	0.168	0.158
		Cor	0.885	0.887	0.882	0.870	0.857
	0.5	Bias	0.041	−0.003	−0.009	−0.040	−0.051
		RMSE	0.154	0.160	0.154	0.164	0.159
		Cor	0.880	0.879	0.878	0.866	0.849

注: CJRT-PINC-θ = 基于能力的联合−交叉负载概率态输入, 噪音连接模型; HO-PINC = 高阶概率态输入, 噪音连接模型; μ_φ = 交叉负载均值; m = 概率态属性; Bias = 所有被试的平均偏差; RMSE = 所有被试的平均均方根误差; Cor = 估计值与真值之间的相关系数。

属性参数估计返真性。首先, CJRT-PINC-θ 在不同模拟条件下属性参数的返真性均较好, 且均优于 HO-PINC 的。题目参数和方差协方差矩阵参数的返真性见网络版附录表 S3.1 和表 S3.2。整体而言, 不同模拟条件下 CJRT-PINC-θ 的参数返真性均较好, 且优于不考虑 RT 的 HO-PINC 的。

4.2 子研究 2: CJRT-PINC-*m*

4.2.1 模型建构

基于图 2(c), CJRT-PINC-*m* 假设被试对属性的掌握情况会影响其完成该题目的耗时, 被试对题目所考查的所有属性的掌握概率越高则对 RT 的影响越大, 即该模型释放了 JRT-PINC 中属性和 RT 之间的条件独立性假设(图 1 中 c)。CJRT-PINC-*m* 同样包含两层模型, 其中 RA 的测量模型为 HO-PINC 模型(公式 3), RT 的测量模型为:

$$logT_{ni} \sim N\left(\xi_i - \tau_n + \kappa_i \prod_{k=1}^{K} m_{nk}^{q_{ik}}, \sigma_i^2\right), \tag{10}$$

式中, κ_i 为交叉负载, 表示理想作答概率对 $\log T_{ni}$ 的影响大小: 当 $\kappa_i > 0$ 时, 理想作答概率越大则 RT 越长, 当 $\kappa_i < 0$ 时, 理想作答概率越大则 RT 越短。能力与加工速度仍为独立分布, 见公式 8。各题目参数也满足多元正态分布:

$$\begin{pmatrix} g_i^{\sim} \\ s_i^{\sim} \\ \xi_i \\ \kappa_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{g^{\sim}} \\ \mu_{s^{\sim}} \\ \mu_{\xi} \\ \mu_{\kappa} \end{pmatrix}, \boldsymbol{\Sigma}_{item}\right), \boldsymbol{\Sigma}_{item} = \begin{pmatrix} \sigma_{g^{\sim}}^2 & \cdots & \\ \vdots & \ddots & \vdots \\ \sigma_{\kappa g^{\sim}} & \cdots & \sigma_{\kappa}^2 \end{pmatrix}, \tag{11}$$

式中各参数含义同上。仍采用贝叶斯 MCMC 算法对 CJRT-PINC-*m* 进行参数估计, 各待估计参数的先验分布见网络版附录 S1。

4.2.2 模拟研究: 数据生成与结果

该部分具体内容见网络版附录 S3.2。整体而言, CJRT-PINC-*m* 在不同模拟条件下模型参数的返真性良好, 均优于不考虑 RT 的 HO-PINC 的。

5 研究 3: 新模型之间的交叉比较

5.1 数据生成与分析

詹沛达(2022)指出在确定态联合 CDM 中"忽略交叉负载所导致的负面结果比冗余考虑交叉负载所导致的更严重。"为验证该结论是否适用于联合概率态 CDM, 研究 3 交叉比较了 JRT-PINC 和两个 CJRT-PINC 的表现。当 JRT-PINC 作为数据生成模型时, 设定能力与加工速度的相关系数($\rho_{\theta\tau}$)为 0.5; 当 CJRT-PINC (CJRT-PINC-θ 和 CJRT-PINC-*m*)作为数据生成模型时, 设定交叉负载(φ_i 和 κ_i)服从均值为 0.5、方差为 0.15 的正态分布。除此之外, 固定样本量 $N = 200$, 题目数量 $I = 15$, 其余参数设定与上文模拟研究相同。每种模拟条件下各生成 30 组数据。评价指标除 RMSE, Bias 和 Cor 外, 还包括模型−数据相对拟合指标 DIC 统计量和模型−数据绝对拟合指标后验预测模型检验(posterior predictive model checking, PPMC)。其中, RA 的 PPMC 采用

$$D(X_{ni}) = \sum_{n=1}^{N}\sum_{i=1}^{I}\left(\frac{X_{ni} - Prob(X_{ni}=1)}{Prob(X_{ni}=1)(1-Prob(X_{ni}=1))}\right)^2,$$

式中 X_{ni} 为 RA, $Prob(X_{ni}=1)$ 为 HO-PINC 模型(公式 3); RT 的 PPMC 采用 $D(logT_{ni}) = \sum_{n=1}^{N}\sum_{i=1}^{I}\left(\frac{logT_{ni} - (\xi_i - \tau_n)}{\sigma_i}\right)^2$, 式中 $logT_{ni}$ 为对数 RT。后验预测概率(posterior predictive probability, *ppp*)接近 0.5 表明模型与数据拟合, 当 *ppp* 值接近 0 或 1 时(如, *ppp* < 0.025 或 *ppp* > 0.975)表明模型与数据不拟合(Bolsinova & Tijmstra, 2018)。

5.2 结果

结果显示所有参数的 PSRF 均小于 1.2, 表示各参数均已收敛。表 5 呈现了各模型在生成数据上的拟合指标。根据 *ppp* 值, 各模型均拟合生成数据, 即联合模型中结构模型的有偏设定不太影响测量模型对数据的绝对拟合。根据 DIC 值发现, 当 JRT-PINC 作为数据生成模型时, 其 DIC 与两个 CJRT-PINC 的 DIC 均接近; 而当任一 CJRT-PINC 作为数据生成模型时, CJRT-PINC 对数据的拟合明显优于 JRT-PINC 的, 这表明引入交叉负载的联合模型的适用范围更广。另外, 当 CJRT-PINC-θ 作为数据生成模型时, 两个 CJRT-PINC 的 DIC 与 JRT-PINC 的 DIC 之间的差值在 1,000 左右; 而当 CJRT-PINC-*m* 作为数据生成模型时, 两个 CJRT-PINC 的 DIC 与 JRT-PINC 的 DIC 之间的差值仅为 100 左右。该结果表明 CJRT-PINC-θ 的普适性相对更高: 当其他两个模型为数据生成模型时, 它能够提供与数据生成模型几乎一致的拟合指标; 而当它作为数据生成模型时, 其他两个模型对数据的拟合则相对要差一些。另外, 各模型参数的返真性也呈现类似的趋势(见网络版附录表 S4.1~S4.4)。总之, 研究 3 结果表明(1)忽略可能存在的交叉负载所导致的负面结果比冗余考虑存在交叉负载所导致的更严重, 即 CJRT-PINC 对测验情境的兼容性优于 JRT-PINC; 且(2) CJRT-PINC-θ 的普适性相对于 CJRT-PINC-*m* 更高。

表 5　研究 3 中模型−数据相对拟合情况

数据生成模型	数据分析模型	DIC	*ppp*.RA	*ppp*.RT
JRT-PINC	JRT-PINC	7935.189	0.479	0.790
	CJRT-PINC-θ	7935.783	0.499	0.792
	CJRT-PINC-*m*	7947.637	0.502	0.792
CJRT-PINC-θ	JRT-PINC	8979.986	0.505	0.706
	CJRT-PINC-θ	7978.639	0.517	0.790
	CJRT-PINC-*m*	8087.316	0.433	0.784
CJRT-PINC-*m*	JRT-PINC	8046.958	0.485	0.781
	CJRT-PINC-θ	7952.253	0.503	0.782
	CJRT-PINC-*m*	7952.049	0.506	0.791

注：DIC = deviance information criterion; *ppp*.RA = RA 的后验预测概率; *ppp*.RT = RT 的后验预测概率。

6　实证数据分析

6.1　数据描述

本节以 PISA 2012 年计算机化数学测验数据为例进一步展现新模型在实践中的可应用性。测验数据共包含 32 个国家，研究从中选取了 4 个国家/地区：中国上海(QCN)、美国(USA)、新加坡(SGP)和斯洛伐克共和国(SVK)。初始样本量为 1754 人，清理后数据包含 $N = 1597$ 名被试。PISA 2012 的数学测验框架(OECD, 2013)已公开的 10 道题目共包含 7 个属性(Zhan, Jiao, & Liao, 2018)，分别是：(K_1)变化和关系、(K_2)数量、(K_3)空间与图形、(K_4)不确定性和数据、(K_5)与职业相关的背景、(K_6)与社会相关的背景以及(K_7)与科学相关的背景。测验 Q 矩阵见网络版附录表 S5.1。另外，CM015Q02D、CM015Q03D 和 CM020Q01 三道题目原为多级评分题目(0, 1, 2)，由于本研究的模型仅针对二级评分数据(0, 1)，因此对这三题的作答结果采用 Zhan, Jiao 和 Liao (2018)的编码方式：0 和 1 编码为 0，2 编码为 1。

6.2　分析与评价指标

分别使用 JRT-PINC、CJRT-PINC-θ、CJRT-PINC-*m* 和另外两个已有模型分析该数据——包括仅可分析 RA 的 HO-PINC 和基于确定态属性的联合作答与时间 DINA (joint responses and times DINA, JRT-DINA; Zhan, Jiao, & Liao, 2018)。采用 DIC 统计量作为模型−数据相对拟合指标，*ppp* 作为模型−数据绝对拟合指标。

6.3　结果

表 6 呈现了各模型在测验数据上的拟合指标。需要注意的是，由于 CJRT-PINC-*m* 中有约 42.75%的待估计参数(包括 m_{nk}, s_i, g_i, κ_i, λ_{0k}, λ_{1k}, ξ_i, $\sigma_{\beta\delta}$, $\sigma_{\beta\kappa}$, $\sigma_{\delta\kappa}$)未达到收敛标准(PSRF < 1.2)，因此该模型与数据的拟合结果仅供参考，后续不在文中进行探讨。由结果可知，所有模型的 *ppp* 值均在 0.5 左右，表示各模型均拟合测验数据。根据 DIC 指标，基于概率态属性的三个联合模型对数据的拟合均优于基于确定态属性的 JRT-DINA 的。同时，基于联合−交叉负载建模框架的两个 CJRT-PINC 对数据的拟合优于基于联合−层级建模框架的 JRT-PINC。下文将主要基于对数据拟合相对最好的 CJRT-PINC-θ 的分析结果进行阐述。

网络版附录中图 S5.1 和图 S5.2 分别呈现了各模型加工速度参数和能力参数估计值的散点图。结果显示，各模型估计的能力参数具有较高的一致性(各模型的估计值间相关系数最小值为 0.853)，同样估计的加工速度参数也具有较高的一致性(各模型的估计值间相关系数最小值为 0.907)；这种高相

表 6　实证数据中模型−数据拟合指标

拟合指标	数据分析模型				
	HO-PINC	JRT-DINA	JRT-PINC	CJRT-PINC-θ	CJRT-PINC-*m*
DIC	17410	50090	41772	41186	40746
ppp.RA	0.728	0.611	0.571	0.638	0.404
ppp.RT	—	0.594	0.596	0.604	0.603

注：DIC = deviance information criterion; *ppp*.RA = RA 的后验预测概率; *ppp*.RT = RT 的后验预测概率; HO-PINC 分析的数据量少于另外 4 个联合模型，所以其 DIC 值不具有可比性。

关说明各模型所定义的能力参数和加工速度参数是大致相同的，测量的是相同的心理特质。另外，各模型的估计值有一定差异性，尤其是能力参数的估计值，这点从模拟研究中能力参数的估计返真性结果也可以反映出来(即两种联合建模法的主要差异体现在能力参数的估计精度上)。以 CJRT-PINC-θ 为参照基准，对加工速度而言，可发现(1)JRT-DINA 和 JRT-PINC 均略低估了高速度区间($\hat{\tau}_n > 1$)内被试的加工速度水平；(2)两个 CJRT-PINC 的估计结果有较高的一致性；对能力而言，可发现(1)HO-PINC 会高估低能力区间($\hat{\theta}_n < -0.5$)内的被试；(2)JRT-DINA 会低估高能力区间($\hat{\theta}_n > 1$)内的被试；(3)两个 CJRT-PINC 的估计结果有较高的一致性；(4)对于少部分在低能力区($\hat{\theta}_n < -1$)间内的被试，基于联合−层级建模的两个模型与基于联合−交叉负载建模的两个模型之间的估计值差异较大。

表 7 呈现了 CJRT-PINC-θ 在测验数据中交叉负载 φ_i 的后验均值和置信区间，以及其他题目参数的后验均值。结果显示交叉负载 φ_i 后验均值的估计范围在−0.027 到 0.338 之间，所有题目上的交叉负载 95%置信区间都不包含零。题目 1 和 9 的 $\varphi_i < 0$，表明这两个题目上能力越高的被试作答题目的时间越短，其他题目则相反($\varphi_i > 0$)。该结果整体与 JRT-PINC 的基本一致，JRT-PINC 中能力与加工速度呈负相关($\rho_{\theta\tau} = -0.531$)，表示能力越高的被试加工速度越慢(作答题目的时间越长)。能力与加工速度负相关结果可能是因为 PISA 属于低风险/低动机的测验，其结果对于学生而言影响较小因此学生的作答动机较低，这一结论与已有研究(Zhan, Jiao, & Liao, 2018)的结果一致。此外 CJRT-PINC-θ 中各题目的交叉负载估计值之间并不一致，表明各题目的 RT 为能力提供的辅助信息量具有差异性，单凭 JRT-PINC 中的一个笼统相关系数可能无法较好地处理该情况。另外，我们计算了交叉负载和其他 3 个题目参数之间的相关系数，发现交叉负载与失误参数的相关系数最高(0.923)，与时间强度参数的相关系数次之(0.622)，与猜测参数的相关系数最低(−0.599)；结果表明题目的失误参数越高，RT 为能力提供的辅助信息越高；当然，这种关系只是该数据特有的，结论的推广性有待进一步验证。

表 8 呈现了 JRT-PINC, CJRT-PINC-θ, JRT-DINA 和 HO-PINC 对个体的分析结果。4 个模型的诊断结

表 7 实证数据中 CJRT-PINC-θ 的交叉负载后验均值和可信区间及其他题目参数估计值

题目	参数	后验均值	95% CI	时间强度参数	失误参数	猜测参数
CM015Q01	φ_1	−0.027	[−0.028, −0.025]	4.228	0.022	0.353
CM015Q02D	φ_2	0.338	[0.337, 0.340]	4.616	0.135	0.001
CM015Q03D	φ_3	0.321	[0.319, 0.323]	4.686	0.092	0.008
CM020Q01	φ_4	0.066	[0.065, 0.068]	4.812	0.054	0.044
CM020Q02	φ_5	0.050	[0.049, 0.052]	3.864	0.030	0.364
CM020Q03	φ_6	0.106	[0.104, 0.107]	4.314	0.039	0.137
CM020Q04	φ_7	−0.009	[−0.011, −0.008]	3.746	0.036	0.279
CM038Q03T	φ_8	0.168	[0.166, 0.169]	4.226	0.075	0.439
CM038Q05	φ_9	0.193	[0.192, 0.195]	4.572	0.064	0.048
CM038Q06	φ_{10}	0.137	[0.135, 0.139]	4.462	0.054	0.021

表 8 实证数据中各模型对个体属性的诊断结果示例

被试	诊断模型	K_1	K_2	K_3	K_4	K_5	K_6	K_7	θ	τ
59	HO-PINC	0.071	0.731	0.337	0.529	0.498	0.459	0.337	−0.193	–
	JRT-PINC	0.483	0.506	0.353	0.855	0.359	0.409	0.593	−0.213	0.401
	CJRT-PINC-θ	0.656	0.623	0.273	0.553	0.493	0.746	0.493	−0.019	0.349
	JRT-DINA	0	1	1	1	0	1	0	−0.040	0.481
299	HO-PINC	0.667	0.977	0.770	0.846	0.967	0.801	0.780	1.051	–
	JRT-PINC	0.711	0.907	0.587	0.968	0.870	0.555	0.790	1.058	0.325
	CJRT-PINC-θ	0.850	0.928	0.558	0.681	0.901	0.867	0.739	0.975	0.283
	JRT-DINA	1	1	1	1	1	1	0	0.784	0.478
977	HO-PINC	0.667	0.976	0.771	0.844	0.965	0.798	0.778	1.059	–
	JRT-PINC	0.709	0.906	0.588	0.968	0.871	0.556	0.794	1.063	0.327
	CJRT-PINC-θ	0.841	0.918	0.537	0.674	0.888	0.862	0.725	0.914	0.285
	JRT-DINA	1	1	1	1	1	1	1	1.162	0.664

构具有一定的一致性, 但同时存在差异。首先, 当概率态属性估计结果大于 0.5 时, 确定态属性的诊断结果也多为“1”。其次, 基于概率态属性的两个模型比基于确定态属性的 JRT-DINA 能更精细地反馈出学生对属性的掌握情况及个体差异性。以被试 59 为例, JRT-DINA 诊断被试的属性 2 为掌握“1”, 但此时概率态属性结果显示被试对属性 2 的掌握概率仅略高于 0.5, 距离熟练/完全掌握还有一定距离, 还需要进一步干预。再次, 与 HO-PINC 相比, 三个考虑 RT 信息的联合模型均可以提供有关加工速度的估计值, 丰富了反馈报告的内容。

7　总结、讨论与展望

7.1　总结与讨论

随着实践者对精细化诊断需求的不断增加, 传统基于确定态属性的 CDM 已显得力不从心, 概率态 CDM 应运而生。但是已有的概率态 CDM 仍然延续传统的建模方式, 无法实现同时分析多模态数据(比如, RA 和 RT)。对此, 本文基于两种可联合分析 RA 和 RT 的认知诊断建模框架提出了三个联合概率态 CDM：JRT-PINC、CJRT-PINC-θ 和 CJRT-PINC-*m*。其中, JRT-PINC 是基于联合−层级认知诊断建模框架构建的, 但该模型需要满足较多的条件独立性假设以保证模型的可识别性; 另外两个 CJRT-PINC 是基于联合−交叉负载认知诊断建模框架构建的, 它们释放了 JRT-PINC 中部分条件独立性假设, 增加了适用范围。三个新模型均具有概率态 CDM 的优点, 可以实现对属性掌握状态的精细化诊断, 有益于区分被试间的个体差异, 为因材施教和精准教学提供数据支持; 同时, 作为一种联合模型, 三者均可利用 RT 中所包含的信息提高模型参数估计精度, 并反馈个体加工速度, 丰富诊断反馈信息。

本文通过三个模拟研究探讨了新模型在不同模拟条件下的心理计量学性能。模拟研究结果主要表明：(1)三个新模型的参数估计返真性均较好; (2)额外引入 RT, 不仅有助于提高模型参数估计精度还可提供有关个体加工速度的测量; (3) CJRT-PINC-θ 比 CJRT-PINC-*m* 更充分地利用 RT 所包含信息去提高核心建构(能力和属性)的参数估计精度; (4)忽略交叉负载所导致的负面结果比冗余考虑交叉负载所导致的更严重。然后, 本文以一则实证数据为例对比探究了 5 个 CDM 的表现, 包括 3 个联合概率态模型(JRT-PINC、CJRT-PINC-θ、CJRT-PINC-*m*)、1 个联合确定态模型(JRT-DINA)和 1 个仅分析 RA 的概率态模型(HO-PINC)。研究结果表明(1)相比于确定态属性, 概率态属性可以更精细化地反馈被试对属性的掌握情况; (2)基于联合−交叉负载认知诊断建模框架构建的 CJRT-PINC 比基于联合−层级认知诊断建模框架构建的 JRT-PINC 更拟合该数据; (3)额外引入 RT, 可丰富诊断反馈内容, 提供有关被试加工速度的测量。

综上, 新模型的提出丰富了概率态 CDM 的适用范围, 为后续进一步在技术增强型测评系统中联合分析多模态数据进行精细化全面化诊断提供了方法学引导。当然, 本研究仅在有限的范围内探讨并展现了三个联合概率态 CDM 的表现。由于任何模型都有其适用的测验情境, 本研究并没有否定其他对比模型, 更多地是进一步丰富现有的认知诊断模型可选项。在实践中, 可使用数据驱动方法, 依据模型−数据相对拟合指标来选择合适的模型。

7.2　局限与展望

因能力、精力和篇幅所限, 本文还存在一些局限性有待后续研究探讨。首先, 本文仅以一些相对常用但简约的模型为例来说明建模思路。比如, 针对 RA 仅选用了满足连接缩合规则的 HO-PINC 模型; 针对 RT 仅选用了假设 RT 为对数正态分布的 LRTM。这导致文中所提模型仅能处理有限的测验情境(如, 二级评分 RA、加工速度恒定、被试群体具有同质性、不涉及协变量信息)。当然, 鉴于新建模框架的灵活性, 后续研究可尝试替换不同测量模型以满足不同的测量需求。其次, 本研究遵循联合−层级建模框架和联合−交叉负载建模框架, 假设变量之间为线性关系(如, JRT-PINC 中能力和加工速度之间, CJRT-PINC 中能力或属性和 RT 之间)。尽管本研究模型−数据拟合指标表明包含该线性假设的模型可以拟合数据, 后续研究仍可从更严谨的角度检验该线性假设是否成立; 比如, 可以通过验证能力估计值($\hat{\theta}$)与速度估计值解释后 RT 残差($logT_{ni}-(\hat{\xi}_i-\hat{\tau}_n)$)之间是否满足线性关系来验证该假设(Bolsinova & Tijmstra, 2018)。然后, 模拟研究中操纵变量的数量及变量所包含的水平数量均有限, 尚未充分反映新模型的心理计量学性能。后续研究可引入其他操纵变量(如, 属性之间的相关性、Q 矩阵的错误设定、不同的属性截距和区分度参数、数据的缺失值比例及异常作答)从不同视角进一步探讨新模型的心理计量学性能。

《深化新时代教育评价改革总体方案》中提出

“改进结果评价，强化过程评价，探索增值评价，健全综合评价”；在“改进结果评价”的基础上，未来可面向过程评价、增值评价和综合评价，进一步尝试拓展本文提出的模型，比如：

(1) “强化过程评价”取向。概率态属性的一个有价值的实践用途是描述被试对属性的掌握状况的精细化变化(Zhan, 2021)。本研究暂局限于横断测验，未关注被试潜在特质的动态发展变化。未来可尝试结合已有的纵向认知诊断建模，进一步探索概率态属性和加工速度的动态变化。

(2) “探索增值评价”取向。CDM 主要关注对个体水平特质的测量，暂未关注对教师和学校层面的投入的评价。结合“过程评价”取向的未来成果，精细化诊断适宜于刻画学生的发展变化，不仅有助于肯定学生的努力付出，也有助于准确反映教师和学校的投入，进而有利于实现基于学生学习进步来评价教师教学绩效的主张(张莉娜 等, 2022)。

(3) “健全综合评价”取向。本研究对多模态数据的利用尚不够充分，局限于 RA 和 RT 两种数据，进而仅能提供与认知能力和加工速度有关的诊断反馈。随着信息技术及测量方式的发展，技术增强型测评已经可以获取学生问题解决过程中的多模态数据(如，眼动数据、鼠标点击次数、行动序列，以及面部表情、动作和心率等变化数据)，后续可以将更多模态的数据纳入分析中(如，Zhan et al., 2022)，以丰富数据分析结果所包含的信息，为全面化、多元化的综合评价提供方法学支持。

参 考 文 献

Bolsinova, M., de Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika, 82*, 1126–1148.

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more from response times. *British Journal of Mathematical and Statistical Psychology, 71*(1), 13–38.

Bradshaw, L., & Levy, R. (2019). Interpreting probabilistic classifications from diagnostic psychometric models. *Educational Measurement: Issues and Practice, 38*(2), 79–88.

de Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology, 10*, 102.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*, 333–353.

Guo, L. Shang, P., & Xia, L. (2017). Advantages and illustrations of application of response time model in psychological and educational testing. *Advances in Psychological Science, 25*(4), 701–712.

[郭磊, 尚鹏丽, 夏凌翔. (2017). 心理与教育测验中反应时模型应用的优势与举例. *心理科学进展, 25*(4), 701–712.]

Huang, H.-Y. (2020). Utilizing response times in cognitive diagnostic computerized adaptive testing under the higher-order deterministic input, noisy ‘and’ gate model. *British Journal of Mathematical and Statistical Psychology, 73*(1), 109–141.

Liu, Q., Wu, R. Z., Chen, E. H., Xu, G. D., Su, Y., Chen, Z. G., & Hu, G. P. (2018). Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Transactions on Intelligent Systems and Technology, 9*(4), Article 48.

Mao, X. (2014). The attribute mastery probability cognitive diagnostic model. *Journal of Sichuan Normal University (National Science), 37*(3), 437–443.

[毛秀珍. (2014). 基于属性掌握概率的认知诊断模型. *四川师范大学学报(自然科学版), 37*(3), 437–443.]

Meng, X., Tao, J., & Chang, H.-H. (2015). A conditional joint modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement, 52*(1), 1–27.

OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. OECD Publishing.

Peng, S., Cai, Y., Wang, D., Luo, F., & Tu, D. (2022). A generalized diagnostic classification modeling framework integrating differential speediness: Advantages and illustrations in psychological and educational testing. *Multivariate Behavioral Research, 57*(6), 940–959.

Ranger, J. (2013). A note on the hierarchical model for responses and response times in tests of van der Linden (2007). *Psychometrika, 78*(3), 538−544

Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal cognitive diagnostic assessment. *AERA Open, 7*(1), 1−15.

Tatsuoka, K. K. (1983). Rule Space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*(2), 181–204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*, 287–308.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology, 68*(3), 456–477.

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika, 81*, 625–649.

Yan, J. H. (2010). Cognitive styles affect choice response time and accuracy. *Personality and Individual Differences, 48*, 747–751.

Zhan, P. (2021). Refined learning tracking with a longitudinal probabilistic diagnostic model. *Educational Measurement: Issues and Practice, 40*(1), 44–58.

Zhan, P. (2022). Joint-cross-loading multimodal cognitive

diagnostic modeling incorporating visual fixation counts. *Acta Psychologica Sinica, 54*(11), 1416–1423.
[詹沛达. (2022). 引入眼动注视点的联合-交叉负载多模态认知诊断建模. *心理学报, 54*(11), 1416–1423.]
Zhan, P., & Bian, Y. (2015). The probabilistic-inputs, noisy “and” gate model. *Journal of Psychological Science, 38*(5), 1230–1238.
[詹沛达, 边玉芳. (2015). 概率性输入, 噪音“与”门(PINA)模型. *心理科学, 38*(5), 1230–1238.]
Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology, 71*(2), 262–286.
Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics, 44*(4), 473–503.
Zhan, P., Man, K., Wind, S. A., & Malone, J. (2022). Cognitive diagnosis modelling incorporating response times and fixation counts: Providing comprehensive feedback and accurate diagnosis. *Journal of Educational and Behavioral Statistic, 47*(6), 736rnal.
Zhan, P., Tian, Y., Yu, Z., Li, F., & Wang, L. (2020). A comparative study of probabilistic logic and fuzzy logic in refined learning diagnosis. *Journal of Psychological Science, 43*, 1258–1266.
[詹沛达, 田亚淑, 于照辉, 李菲茗, 王立君. (2020). 概率逻辑与模糊逻辑在精细化学习诊断中的对比研究. *心理科学, 43*, 1258–1266.]
Zhan, P., Wang, W. C., Jiao, H., & Bian, Y. F. (2018). Probabilistic-input, noisy conjunctive models for cognitive diagnosis. *Frontiers in Psychology, 9*, 997.
Zhang, L., Zhong, Z., Liu, H., & You, X. (2022). Exploration and reflection on teachers' performance appraisal in the context of educational evaluation reform-based on the perspective of value-added evaluation. *Educational Science Research, 8*, 23–39.
[张莉娜, 钟祖荣, 刘红云, 游晓锋. (2022). 教育评价改革背景下教师绩效考评的探索与思考——基于增值评价的视角. *教育科学研究, 8*, 23–39.]
Zheng, T., Zhou, W., & Guo, L. (2023). Cognitive diagnosis modelling based on response times. *Journal of Psychological Science, 46*(2), 478–490.
[郑天鹏, 周文杰, 郭磊. (2023). 基于题目作答时间信息的认知诊断模型. *心理科学, 46*(2), 478–490.]

Joint cognitive diagnostic modeling for probabilistic attributes incorporating item responses and response times

TIAN Yashu, ZHAN Peida, WANG Lijun

(*School of Psychology, Zhejiang Normal University; Intelligent Laboratory of Child and Adolescent Mental Health and Crisis Intervention of Zhejiang Province; Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Jinhua* 321004, *China*)

Abstract

Compared with the conventional CDM with deterministic or binary attributes, the CDM with probabilistic attributes (probabilistic-CDM) can achieve a more refined diagnosis of attribute mastery status, which helps distinguish individual differences between students and provides more reference information for teacher feedback. However, existing probabilistic CDMs can only analyze a single modal of data—item response accuracy (RA), ignoring other modals of data such as item response times (RTs). RTs reflect the cognitive processing speed of the participant. With the popularity of computerized testing, recording RT data has become routine. However, how to use RTs in probabilistic CDM to further improve parameter estimation accuracy and enrich the diagnostic feedback information is still an unsolved methodological problem. To this end, the current study proposes three joint probabilistic CDMs based on the joint-hierarchical and joint-cross-loading cognitive diagnostic modeling approaches.

First, based on joint-hierarchical modeling, the joint-hierarchical probabilistic CDM (JRT-PINC) was proposed in Study 1, which achieved the purpose of using RT to improve diagnostic accuracy. A simulation study was conducted to investigate the psychometric performance of the JRT-PINC under various simulated testing conditions, in which three independent variables, including sample size, test length, and the correlation between person parameters, were manipulated. Second, two joint-cross-loading probabilistic CDMs (CJRT-PINC-θ and CJRT-PINC-*m*) were proposed based on the joint-cross-loading modeling. In contrast to the JRT-PINC model, two CJRT-PINC models directly used RTs to provide information for latent abilities or

attributes by introducing item-level cross-loading parameters. Two CJRT-PINC models released some conditional independence assumptions in JRT-PINC, increasing their application scope. Two simulation studies were conducted to explore their performance under different simulated conditions with different degrees of cross-loading. Third, Study 3 aims to explore the relative merits of the JRT-PINC and two CJRT-PINC models, that is, the necessity of considering cross-loading in the joint analysis of RA and RT. Finally, an empirical example was conducted to illustrate the practical applicability of the proposed models and to compare them with existing CDMs (e.g., CDMs with deterministic attributes).

The simulation results mainly indicated that: (1) all three proposed models can be well recovered under different simulated conditions; (2) CJRT-PINC-θ makes fuller use of the information contained in RTs and thus improves the accuracy of the parameter estimation of the core constructs (e.g., latent ability and attributes) than CJRT-PINC-*m*; and (3) the adverse effects of ignoring the possible cross-loadings are more severe than redundantly considering them. The results of the empirical example indicated that: (1) probabilistic attributes provide more refined feedback on participants' mastery of attributes than deterministic attributes; and (2) two CJRT-PINC models fit this data better than the JRT-PINC model.

Overall, this paper introduced RTs in probabilistic CDM for the first time and proposed three joint probabilistic CDMs based on two joint cognitive diagnostic modeling approaches. This study enriched the scope of application of probabilistic CDMS and provided methodological guidance for further refined and comprehensive diagnosis by jointly analyzing multi-modal data in technology-enhanced assessment systems.

Keywords cognitive diagnosis, probabilistic attribute, item response time, joint modeling framework, cross loading

附录:

S1　参数估计中的先验分布及鲁棒性分析

S1.1 先验分布

S1.1.1　JRT-PINC

首先, 根据条件独立性假设, X_{ni} 和 $\log T_{ni}$ 满足:

$$X_{ni} \sim Bernoulli(P(X_{ni}=1)), \quad \log T_{ni} \sim N(\xi_i - \tau_n, \sigma_i^2).$$

其次, 假设被试参数先验分布为:

$$\begin{pmatrix}\theta_n \\ \tau_n\end{pmatrix} \sim MVN\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \boldsymbol{\Sigma}_{\text{person}}\right).$$

为了使模型可识别, 约束 $\mu_\theta = 0$, $\mu_\tau = 0$, 以及 $\sigma_\theta = 1$ (Guo et al., 2020; Meng, Tao, & Chang, 2015; van der Linden, 2007; Zhan et al., 2018)。由于矩阵中 $\sigma_\theta = 1$ 不能直接使用逆 Wishart 分布(inverse-Wishart), 因此需要对 Σ_{person} 进行 Cholesky 分解(Zhan et al., 2018), 即:

$$\boldsymbol{\Sigma}_{\text{person}} = \Delta_{\text{person}}\Delta'_{\text{person}}, \quad \Delta_{\text{person}} = \begin{pmatrix}1 & 0 \\ \phi & \psi\end{pmatrix},$$

其中, Δ_{person}是一个下三角矩阵, 由于Cholesky分解的下三角矩阵的对角元是大于零的, 因此 $\psi > 0$; Δ'_{person} 是 Δ_{person} 的共轭转置。Δ_{person} 中参数的超先验(hyper prior)设定为: $\phi \sim N(0,1), \psi \sim \text{Gamma}^+(1,1)$。

此外, 高阶结构参数的先验分布设定为: $\lambda_{0k} \sim N(0, 4)$, $\lambda_{1k} \sim N^+(0, 4)$。

最后, 假设题目参数先验分布为:

$$\begin{pmatrix} logit(g_i) \\ logit(1-s_i) - logti(g_i) \\ \xi_i \end{pmatrix} \equiv \begin{pmatrix} g_i^{\sim} \\ s_i^{\sim} \\ \xi_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{g^\sim} \\ \mu_{s^\sim} \\ \mu_\xi \end{pmatrix}, \boldsymbol{\Sigma}_{\text{item}}\right),$$

其中各参数的超先验设定为:

$$\mu_{g^\sim} \sim N(-2.197, 2), \quad \mu_{s^\sim} \sim N^+(4.394, 2), \quad \mu_\xi \sim N(3, 2), \quad \boldsymbol{\Sigma}_{\text{item}} \sim \text{InvWishart}(R, 3),$$

式中, R 是一个三元单位矩阵(identity matrix)。另外, $\sigma_i^2 \sim Gamma(1,1)$。

S1.1.2　CJRT-PINC

由于 CJRT-PINC-θ 中潜在能力和速度参数为独立分布, 因此假设被试参数先验分布为:

$$\theta_n \sim N(0,1), \quad \tau_n \sim N(0,1).$$

假设题目参数先验分布为:

$$\begin{pmatrix} logit(g_i) \\ logit(1-s_i) - logti(g_i) \\ \xi_i \\ \varphi_i \end{pmatrix} \equiv \begin{pmatrix} g_i^{\sim} \\ s_i^{\sim} \\ \xi_i \\ \varphi_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{g^\sim} \\ \mu_{s^\sim} \\ \mu_\xi \\ \mu_\varphi \end{pmatrix}, \boldsymbol{\Sigma}_{\text{item}}\right),$$

$$1/\sigma_i^2 \sim InvGamma(1,1),$$

式中各参数的超先验设定为: $\mu_\varphi \sim N(0, 0.25)$, $\boldsymbol{\Sigma}_{\text{item}} \sim \text{InvWishart}(R, 4)$。其余参数先验分布与 S1.1.1 一致; CJRT-PINC-*m* 的参数先验分布与 CJRT-PINC-θ 的一致(除交叉负载符号不同, $\mu_\kappa \sim N(0, 0.25)$,)不再赘述。

S1.2　鲁棒性分析

S1.2.1　低信息量先验分布

低信息量先验分布的设定以“无知”为前提, 并以大方差(如, 5)为变异范围。由于 S1.1 中部分参数已经采用低信息量先验或超先验, 所以在 S1.1 的基础上, 部分参数的低信息量先验分布设定如下:

$$\mu_{g^\sim} \sim N(0,5), \quad \mu_{s^\sim} \sim N^+(0,5), \quad \mu_\xi \sim N(0,5), \quad \mu_\varphi \sim N(0,5), \quad \mu_\kappa \sim N(0,5),$$

其他参数的先验分布保持不变。

S1.2.2　参数估计一致性

选用正文模拟研究中 N = 200, I = 15 条件下生成的数据作为分析模型, Q 矩阵见正文图 3; 该模拟条件属于小样本短测验情境, 理论上参数估计结果受到先验分布的影响更大。随样本量增大及测验长度提高, 参数估计结果受先验分布中所含信息量的影响会逐渐降低(即鲁棒性会增加)。另外, 对于 JRT-PINC 模型, 设定潜在能力与加工速度的相关系数 $\rho_{\theta\tau}$ = 0.5; 对于 JRT-PINC-θ, 设定交叉负载(φ_i)满足均值为 0.5、方差为 0.15 的正态分布; 对于, JRT-PINC-m, 设定交叉负载(κ_i)满足均值为 0.5、方差为 0.15 的正态分布。三个模型的其他参数设定与正文研究中保持一致, 且参数估计设定(如, 马尔可夫链长)与各模拟研究中保持一致。

表 S1.1 到表 S1.3 呈现了 3 个模型在不同信息量先验分布下各参数的返真性。整体而言, 基于研究经验设定的正常信息量先验分布下的参数估计返真性与低信息量先验分布下的参数估计返真性具有较高的一致性, 且前者略优于后者。另外, 三个模型中, CJRT-PINC-θ 受先验分布信息量的影响略大于另外两个模型。

表 S1.1　不同信息量先验分布下被试参数的估计一致性

模型	信息量	θ			τ		
		Bias	RMSE	Cor	Bias	RMSE	Cor
JRT-PINC	正常	−0.000	0.458	0.887	−0.000	0.120	0.902
	低	−0.003	0.459	0.888	−0.002	0.122	0.901
CJRT-PINC-θ	正常	−0.007	0.287	0.958	−0.001	0.194	0.971
	低	−0.008	0.385	0.891	0.000	0.192	0.890
CJRT-PINC-m	正常	−0.018	0.446	0.894	−0.005	0.136	0.985
	低	−0.018	0.448	0.894	−0.006	0.140	0.941

表 S1.2　不同信息量先验分布下属性的估计一致性

模型	信息量	指标	m_1	m_2	m_3	m_4	m_5
JRT-PINC	正常	Bias	0.006	0.008	0.008	−0.008	−0.022
		RMSE	0.128	0.139	0.144	0.147	0.142
		Cor	0.898	0.899	0.894	0.886	0.874
	低	Bias	0.007	0.007	0.004	−0.011	−0.016
		RMSE	0.123	0.137	0.133	0.136	0.134
		Cor	0.890	0.889	0.886	0.879	0.866
CJRT-PINC-θ	正常	Bias	−0.004	−0.030	−0.006	−0.009	−0.002
		RMSE	0.097	0.118	0.107	0.111	0.098
		Cor	0.952	0.949	0.952	0.950	0.949
	低	Bias	0.011	−0.021	−0.002	−0.007	−0.012
		RMSE	0.122	0.136	0.12	0.125	0.112
		Cor	0.923	0.894	0.896	0.884	0.904
CJRT-PINC-m	正常	Bias	0.012	−0.010	0.002	−0.009	−0.005
		RMSE	0.127	0.124	0.132	0.135	0.135
		Cor	0.907	0.910	0.906	0.897	0.882
	低	Bias	0.009	−0.010	−0.001	−0.012	−0.010
		RMSE	0.127	0.124	0.132	0.135	0.136
		Cor	0.907	0.910	0.905	0.897	0.882

表 S1.3　不同信息量先验分布下属性的估计一致性

模型	先验分布	g		s		ξ		$1/\sigma^2$		交叉负载	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
JRT-PINC	正常	0.007	0.055	0.023	0.060	−0.002	0.037	0.011	0.028	–	–
	低	0.014	0.056	0.035	0.067	−0.003	0.038	0.012	0.028	–	–
CJRT-PINC-θ	正常	0.002	0.055	0.002	0.049	0.001	0.061	0.012	0.062	−0.029	0.030
	低	0.033	0.108	0.037	0.114	0.002	0.061	−0.053	0.162	0.012	0.030
CJRT-PINC-*m*	正常	0.002	0.048	0.009	0.054	0.035	0.097	0.010	0.204	−0.102	0.028
	低	0.010	0.049	0.018	0.058	0.020	0.097	−0.069	0.198	0.010	0.028

S2　研究 1 补充结果

表 S2.1　研究 1 中概率态属性参数估计的平均 Bias

N	*I*	$\rho_{\theta\tau}$	JRT-PINC					HO-PINC				
			m_1	m_2	m_3	m_4	m_5	m_1	m_2	m_3	m_4	m_5
200	15	−0.5	0.008	0.016	0.007	−0.003	−0.013	0.049	0.022	0.015	−0.017	−0.048
		−0.3	0.009	0.010	0.005	−0.007	−0.016	0.051	0.015	0.012	−0.022	−0.051
		0	0.008	0.014	0.000	−0.007	−0.018	0.056	0.019	0.009	−0.017	−0.049
		0.3	0.004	0.010	0.006	−0.011	−0.018	0.045	0.022	0.011	−0.017	−0.053
		0.5	0.006	0.008	0.008	−0.008	−0.022	0.051	0.019	0.017	−0.014	−0.059
	30	−0.5	0.002	−0.007	−0.016	−0.002	−0.005	0.004	−0.014	−0.026	−0.029	−0.040
		−0.3	0.004	−0.003	−0.016	−0.006	−0.003	0.006	−0.008	−0.026	−0.033	−0.035
		0	0.004	−0.006	−0.017	−0.008	−0.001	0.005	−0.013	−0.031	−0.034	−0.033
		0.3	0.005	−0.010	−0.013	−0.007	0.001	0.008	−0.018	−0.029	−0.037	−0.032
		0.5	0.004	−0.009	−0.013	−0.004	0.004	0.008	−0.017	−0.028	−0.030	−0.029
500	15	−0.5	−0.008	−0.006	−0.012	0.022	−0.024	0.019	−0.000	−0.020	0.010	−0.054
		−0.3	−0.007	−0.008	−0.008	0.020	−0.027	0.017	−0.001	−0.016	0.013	−0.060
		0	−0.010	−0.007	−0.006	0.014	−0.031	0.018	−0.003	−0.015	0.003	−0.062
		0.3	−0.016	−0.008	−0.004	0.012	−0.029	0.007	−0.001	−0.008	−0.000	−0.060
		0.5	−0.012	−0.006	−0.004	0.012	−0.026	0.007	−0.003	−0.009	0.000	−0.057
	30	−0.5	0.008	0.011	−0.018	0.013	−0.017	0.011	0.005	−0.031	−0.008	−0.047
		−0.3	0.005	0.011	−0.020	0.014	−0.018	0.007	0.006	−0.031	−0.008	−0.046
		0	0.006	0.014	−0.016	0.015	−0.018	0.007	0.012	−0.030	−0.009	−0.047
		0.3	0.007	0.013	−0.016	0.018	−0.020	0.009	0.010	−0.030	−0.007	−0.043
		0.5	0.008	0.014	−0.014	0.017	−0.020	0.008	0.010	−0.027	−0.010	−0.041

注: JRT-PINC = 联合−层级概率态输入, 噪音连接模型; HO-PINC = 高阶概率态输入, 噪音连接模型; θ = 高阶潜在能力; τ = 加工速度; N = 样本量; I = 题目数量; ρ = 能力与速度之间相关。

表 S2.2　研究 1 中题目参数的返真性

N	*I*	$\rho_{\theta\tau}$	JRT-PINC								HO-PINC			
			g		s		ξ		$1/\sigma^2$		g		s	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
200	15	−0.5	0.004	0.055	0.022	0.056	−0.001	0.037	0.011	0.028	0.055	0.077	0.118	0.138
		−0.3	0.005	0.054	0.022	0.057	−0.001	0.037	0.011	0.028	0.057	0.078	0.117	0.137
		0	0.006	0.055	0.022	0.056	−0.001	0.036	0.011	0.028	0.056	0.078	0.119	0.139
		0.3	0.007	0.055	0.022	0.057	−0.002	0.037	0.011	0.028	0.057	0.079	0.118	0.138
		0.5	0.007	0.055	0.023	0.060	−0.002	0.037	0.011	0.028	0.057	0.080	0.120	0.141

续表 S2.2

N	I	$\rho_{\theta\tau}$	JRT-PINC								HO-PINC			
			g		s		ξ		$1/\sigma^2$		g		s	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	30	−0.5	0.005	0.043	0.013	0.051	−0.002	0.034	0.012	0.029	0.043	0.061	0.122	0.143
		−0.3	0.004	0.044	0.013	0.051	−0.003	0.034	0.012	0.029	0.042	0.061	0.124	0.146
		0	0.005	0.044	0.013	0.049	−0.002	0.034	0.012	0.029	0.043	0.062	0.123	0.145
		0.3	0.005	0.043	0.014	0.049	−0.004	0.035	0.012	0.029	0.043	0.062	0.123	0.145
		0.5	0.004	0.043	0.013	0.050	−0.003	0.034	0.012	0.029	0.043	0.061	0.124	0.145
500	15	−0.5	0.006	0.043	0.011	0.044	0.000	0.022	0.004	0.017	0.046	0.061	0.083	0.099
		−0.3	0.006	0.044	0.010	0.043	−0.001	0.022	0.004	0.017	0.047	0.061	0.083	0.099
		0	0.007	0.044	0.009	0.045	−0.000	0.022	0.004	0.017	0.049	0.063	0.084	0.101
		0.3	0.008	0.045	0.010	0.045	−0.001	0.022	0.004	0.017	0.049	0.063	0.082	0.100
		0.5	0.006	0.044	0.008	0.045	−0.000	0.022	0.004	0.017	0.049	0.064	0.083	0.100
	30	−0.5	−0.000	0.029	0.002	0.037	0.000	0.021	0.005	0.017	0.027	0.039	0.080	0.095
		−0.3	0.000	0.029	0.003	0.036	0.000	0.022	0.005	0.017	0.027	0.039	0.080	0.096
		0	−0.001	0.029	0.003	0.037	0.001	0.021	0.005	0.017	0.027	0.039	0.079	0.094
		0.3	−0.001	0.029	0.001	0.035	0.000	0.022	0.005	0.017	0.026	0.038	0.078	0.093
		0.5	−0.002	0.029	0.000	0.035	0.001	0.022	0.005	0.017	0.026	0.038	0.078	0.093

表 S2.3　研究 1 中方差协方差矩阵和题目均值向量的平均 Bias

N	I	$\rho_{\theta\tau}$	Σ_{item}						Σ_{person}		题目均值向量		
			Σ_{11}	Σ_{12}	Σ_{13}	Σ_{22}	Σ_{23}	Σ_{33}	Σ_{12}	Σ_{22}	μ_{g}	μ_{s}	μ_{ξ}
200	15	−0.5	0.000	0.200	0.026	−0.037	−0.053	0.095	−0.003	0.002	0.055	−0.179	−0.001
		−0.3	−0.025	0.226	0.031	−0.065	−0.056	0.095	−0.002	0.002	0.073	−0.194	−0.001
		0	−0.027	0.230	0.030	−0.082	−0.052	0.095	0.001	0.003	0.078	−0.190	−0.012
		0.3	−0.025	0.225	0.032	−0.079	−0.052	0.095	0.003	0.003	0.085	−0.203	−0.002
		0.5	−0.002	0.197	0.026	−0.031	−0.042	0.094	0.002	0.003	0.083	−0.198	−0.002
	30	−0.5	0.033	0.106	−0.006	−0.092	−0.013	0.046	−0.006	0.005	0.050	−0.124	−0.002
		−0.3	0.037	0.103	−0.004	−0.087	−0.017	0.046	−0.007	0.005	0.045	−0.119	−0.003
		0	0.022	0.114	−0.002	−0.107	−0.015	0.046	−0.004	0.005	0.059	−0.142	−0.007
		0.3	0.043	0.088	−0.003	−0.078	−0.012	0.046	−0.002	0.005	0.050	−0.140	−0.004
		0.5	0.044	0.085	−0.003	−0.071	−0.012	0.046	−0.000	0.005	0.041	−0.127	−0.003
500	15	−0.5	0.096	0.053	−0.004	0.087	−0.009	0.096	0.002	0.002	0.053	−0.090	0.000
		−0.3	0.100	0.057	−0.010	0.075	−0.004	0.095	0.002	0.002	0.051	−0.085	−0.001
		0	0.091	0.072	−0.005	0.063	−0.007	0.096	0.002	0.002	0.071	−0.085	−0.011
		0.3	0.079	0.096	−0.007	0.034	−0.009	0.096	0.003	0.002	0.071	−0.091	−0.001
		0.5	0.101	0.078	−0.011	0.048	−0.005	0.096	0.003	0.002	0.047	−0.051	−0.000
	30	−0.5	0.057	−0.019	−0.010	0.071	0.020	0.044	−0.003	0.002	−0.007	0.018	0.000
		−0.3	0.053	−0.013	−0.009	0.063	0.019	0.045	−0.002	0.002	−0.001	−0.001	0.000
		0	0.067	−0.036	−0.013	0.087	0.025	0.044	−0.002	0.002	−0.006	0.004	−0.004
		0.3	0.071	−0.048	−0.012	0.096	0.025	0.044	−0.002	0.002	−0.019	0.024	0.000
		0.5	0.078	−0.058	−0.012	0.108	0.025	0.044	−0.001	0.002	−0.028	0.041	0.001

表 S2.4　研究 1 中方差协方差矩阵和题目均值向量的平均 RMSE

N	I	$\rho_{\theta\tau}$	Σ_{item}						Σ_{person}		题目均值向量		
			Σ_{11}	Σ_{12}	Σ_{13}	Σ_{22}	Σ_{23}	Σ_{33}	Σ_{12}	Σ_{22}	$\mu_{\tilde{g}}$	$\mu_{\tilde{s}}$	μ_{ξ}
200	15	−0.5	0.325	0.438	0.081	0.471	0.122	0.095	0.018	0.007	0.195	0.296	0.012
		−0.3	0.287	0.414	0.077	0.437	0.119	0.096	0.020	0.007	0.195	0.333	0.010
		0	0.277	0.408	0.079	0.431	0.113	0.095	0.019	0.007	0.191	0.337	0.016
		0.3	0.275	0.409	0.074	0.420	0.113	0.095	0.019	0.007	0.213	0.363	0.011
		0.5	0.287	0.430	0.074	0.482	0.118	0.095	0.018	0.007	0.199	0.376	0.011
	30	−0.5	0.176	0.217	0.050	0.269	0.085	0.047	0.014	0.007	0.126	0.265	0.008
		−0.3	0.184	0.218	0.051	0.257	0.089	0.046	0.015	0.008	0.137	0.266	0.007
		0	0.197	0.231	0.048	0.269	0.081	0.047	0.017	0.008	0.145	0.242	0.010
		0.3	0.191	0.218	0.046	0.250	0.081	0.047	0.018	0.008	0.152	0.242	0.009
		0.5	0.192	0.218	0.046	0.248	0.078	0.046	0.017	0.007	0.158	0.259	0.008
500	15	−0.5	0.285	0.325	0.067	0.415	0.089	0.096	0.010	0.004	0.206	0.388	0.005
		−0.3	0.284	0.331	0.066	0.432	0.084	0.095	0.011	0.005	0.191	0.368	0.005
		0	0.267	0.336	0.063	0.444	0.083	0.096	0.011	0.006	0.191	0.352	0.012
		0.3	0.278	0.352	0.066	0.423	0.094	0.096	0.010	0.006	0.212	0.386	0.005
		0.5	0.294	0.365	0.064	0.463	0.089	0.096	0.010	0.006	0.208	0.388	0.005
	30	−0.5	0.163	0.195	0.036	0.290	0.071	0.045	0.008	0.004	0.114	0.261	0.004
		−0.3	0.163	0.189	0.035	0.280	0.069	0.045	0.008	0.004	0.112	0.239	0.004
		0	0.174	0.202	0.036	0.288	0.071	0.044	0.010	0.004	0.120	0.261	0.006
		0.3	0.170	0.197	0.036	0.295	0.071	0.045	0.010	0.004	0.123	0.248	0.005
		0.5	0.173	0.202	0.037	0.311	0.072	0.044	0.008	0.003	0.124	0.235	0.005

S3　研究 2 补充内容

S3.1　研究 2 (子研究 1)补充结果

表 S3.1　研究 2 (子研究 1)中题目参数的返真性

分析模型	μ_{φ}	指标	g	s	ξ	φ	$1/\sigma^2$
CJRT-PINC-θ	0.1	平均 Bias	0.005	0.011	0.000	−0.007	0.010
		平均 RMSE	0.053	0.055	0.049	0.046	0.028
	0.5	平均 Bias	0.002	0.002	0.001	−0.029	0.012
		平均 RMSE	0.055	0.049	0.061	0.062	0.030
HO-PINC	0.1	平均 Bias	0.061	0.113			
		平均 RMSE	0.081	0.134			
	0.5	平均 Bias	0.065	0.111			
		平均 RMSE	0.084	0.131			

表 S3.2　研究 2 (子研究 1)中题目参数方差协方差矩阵和均值向量的返真性

μ_{φ}	指标	Σ_{item}						题目均值向量		
		Σ_{11}	Σ_{12}	Σ_{13}	Σ_{22}	Σ_{23}	Σ_{33}	$\mu_{\tilde{g}}$	$\mu_{\tilde{s}}$	μ_{ζ}
0.1	平均 Bias	−0.054	0.241	0.025	−0.124	−0.042	0.094	0.081	−0.093	0.000
	平均 RMSE	0.310	0.406	0.080	0.367	0.121	0.095	0.228	0.395	0.027
0.5	平均 Bias	−0.023	0.172	0.009	0.058	−0.010	0.094	0.042	0.043	0.001
	平均 RMSE	0.252	0.374	0.078	0.472	0.123	0.094	0.216	0.411	0.045

S3.2 研究 2 (子研究 2)数据生成与结果

为了更符合实际测试的复杂情境，研究设置不同题目中理想作答概率对 RT 的影响不同，因此数据生成中设定自变量交叉负载(κ_i)满足方差为 0.15 的正态分布，其中低影响效应 $\mu_\kappa = 0.1$，高影响效应 $\mu_\kappa = 0.5$。其余参数设定与研究 1 和研究 2 (子研究 1)相同。根据 CJRT-PINC-*m* (正文公式 2~3 和 10~11)生成所有被试在所有题目上的 RA 和 RT。

使用 CJRT-PINC-*m* 和 HO-PINC 分析该数据。结果显示所有参数的 PSRF 均小于 1.2，表示各参数均已收敛。附录表 S3.3 呈现了能力参数与加工速度参数的返真性，附录表 S3.4 呈现了属性参数的返真性。首先，CJRT-PINC-*m* 对所有参数的返真性均优于 HO-PINC 的。其次，在 CJRT-PINC-*m* 中，由于 RT 没有为能力直接提供信息，所以能力参数的 RMSE 与研究 1 中 JRT-PINC 的基本一致。再有，随着交叉负载均值提高，能力、加工速度和属性的返真性均有所提高。题目参数和方差协方差矩阵参数的返真性见附录表 S3.5 和表 S3.6。整体而言，CJRT-PINC-*m* 在不同模拟条件下模型参数的返真性良好，均优于不考虑 RT 的 HO-PINC 的。

表 S3.3 研究 2 (子研究 2)中被试参数估计返真性

分析模型	μ_φ	θ			τ		
		Bias	RMSE	Cor	Bias	RMSE	Cor
CJRT-PINC-*m*	0.1	−0.007	0.461	0.887	0.000	0.135	0.978
	0.5	−0.018	0.446	0.894	−0.005	0.136	0.985
HO-PINC	0.1	−0.009	0.482	0.876			
	0.5	−0.020	0.480	0.875			

注: CJRT-PINC-*m* = 基于属性的联合−交叉负载概率态输入，噪音连接模型; HO-PINC = 高阶概率态输入，噪音连接模型; θ = 能力; τ = 加工速度; μ_φ = 交叉负载均值; Bias = 所有被试的平均偏差; RMSE = 所有被试的平均均方根误差; Cor = 估计值与真值之间的相关系数。

表 S3.4 研究 2 (子研究 2)概率态属性参数估计返真性

分析模型	μ_φ	指标	m_1	m_2	m_3	m_4	m_5
CJRT-PINC-*m*	0.1	Bias	0.006	−0.014	−0.021	0.012	−0.026
		RMSE	0.129	0.131	0.140	0.154	0.144
		Cor	0.904	0.905	0.901	0.888	0.873
	0.5	Bias	0.012	−0.010	0.002	−0.009	−0.005
		RMSE	0.127	0.124	0.132	0.135	0.135
		Cor	0.907	0.910	0.906	0.897	0.882
HO-PINC	0.1	Bias	0.040	0.016	−0.006	−0.009	−0.054
		RMSE	0.153	0.145	0.149	0.154	0.159
		Cor	0.891	0.894	0.888	0.877	0.861
	0.5	Bias	0.038	0.012	−0.006	−0.021	−0.065
		RMSE	0.154	0.148	0.154	0.156	0.159
		Cor	0.889	0.892	0.886	0.876	0.859

注: CJRT-PINC-*m* = 基于属性的联合−交叉负载概率态输入，噪音连接模型; HO-PINC = 高阶概率态输入，噪音连接模型; μ_φ = 交叉负载均值; *m* = 概率态属性; Bias = 所有被试的平均偏差; RMSE = 所有被试的平均均方根误差; Cor = 估计值与真值之间的相关系数。

表 S3.5 研究 2 (子研究 2)中题目参数的返真性

分析模型	μ_φ	指标	g	s	ξ	κ	$1/\sigma^2$
CJRT-PINC-θ	0.1	平均 Bias	−0.005	−0.001	0.026	−0.071	0.010
		平均 RMSE	0.058	0.050	0.103	0.201	0.029
	0.5	平均 Bias	0.002	0.009	0.035	−0.102	0.010
		平均 RMSE	0.048	0.054	0.097	0.204	0.028
HO-PINC	0.1	平均 Bias	0.057	0.114			
		平均 RMSE	0.078	0.134			
	0.5	平均 Bias	0.064	0.110			
		平均 RMSE	0.082	0.132			

表 S3.6　研究 2 (子研究 2)中题目参数方差协方差矩阵和均值向量的返真性

μ_α	指标	Σ_{item}						题目均值向量		
		Σ_{11}	Σ_{12}	Σ_{13}	Σ_{22}	Σ_{23}	Σ_{33}	$\mu_{\tilde{g}}$	$\mu_{\tilde{s}}$	μ_ζ
0.1	平均 Bias	−0.115	0.319	0.048	−0.167	−0.050	0.087	−0.018	0.150	0.026
	平均 RMSE	0.275	0.429	0.100	0.393	0.112	0.090	0.243	0.431	0.067
0.5	平均 Bias	0.058	0.120	0.014	0.059	−0.031	0.105	0.029	−0.008	0.035
	平均 RMSE	0.363	0.466	0.079	0.670	0.130	0.107	0.170	0.368	0.062

S4　研究 3 补充结果

表 S4.1　研究 3 中被试参数的返真性

数据生成模型	数据分析模型	θ			τ		
		Bias	RMSE	Cor	Bias	RMSE	Cor
JRT-PINC	JRT-PINC	0.000	0.457	0.887	−0.001	0.120	0.951
	CJRT-PINC-θ	−0.002	0.475	0.878	−0.002	0.181	0.885
	CJRT-PINC-*m*	−0.001	0.475	0.877	−0.003	0.132	0.943
CJRT-PINC-θ	JRT-PINC	0.007	0.412	0.910	−0.013	0.496	0.600
	CJRT-PINC-θ	0.009	0.286	0.959	−0.016	0.194	0.886
	CJRT-PINC-*m*	0.010	0.302	0.954	−0.016	0.356	0.740
CJRT-PINC-*m*	JRT-PINC	0.003	0.475	0.879	0.008	0.171	0.902
	CJRT-PINC-θ	0.008	0.466	0.884	0.004	0.143	0.936
	CJRT-PINC-*m*	0.005	0.456	0.889	0.005	0.143	0.937

表 S4.2　研究 3 中属性参数的的平均 Bias

数据生成模型	数据分析模型	m_1	m_2	m_3	m_4	m_5
JRT-PINC	JRT-PINC	0.006	0.007	0.008	−0.008	−0.022
	CJRT-PINC-θ	0.005	0.005	0.007	−0.006	−0.023
	CJRT-PINC-*m*	0.015	0.011	0.010	−0.004	−0.028
CJRT-PINC-θ	JRT-PINC	−0.017	−0.014	−0.019	−0.024	−0.026
	CJRT-PINC-θ	−0.016	−0.009	−0.018	−0.018	−0.017
	CJRT-PINC-*m*	0.016	0.012	0.029	0.020	−0.038
CJRT-PINC-*m*	JRT-PINC	0.031	0.001	−0.020	0.002	−0.015
	CJRT-PINC-θ	0.026	0.009	−0.024	0.002	−0.020
	CJRT-PINC-*m*	0.033	0.011	−0.016	−0.001	−0.013

表 S4.3　研究 3 中属性参数的的平均 RMSE

数据生成模型	数据分析模型	m_1	m_2	m_3	m_4	m_5
JRT-PINC	JRT-PINC	0.126	0.138	0.142	0.147	0.143
	CJRT-PINC-θ	0.128	0.140	0.146	0.148	0.145
	CJRT-PINC-*m*	0.132	0.143	0.144	0.151	0.147
CJRT-PINC-θ	JRT-PINC	0.116	0.131	0.139	0.137	0.133
	CJRT-PINC-θ	0.088	0.103	0.146	0.109	0.103
	CJRT-PINC-*m*	0.156	0.142	0.110	0.130	0.150
CJRT-PINC-*m*	JRT-PINC	0.139	0.150	0.151	0.146	0.143
	CJRT-PINC-θ	0.134	0.149	0.149	0.142	0.141
	CJRT-PINC-*m*	0.137	0.143	0.142	0.139	0.136

表 S4.4　研究 3 中属性参数的的 Cor

数据生成模型	数据分析模型	m_1	m_2	m_3	m_4	m_5
JRT-PINC	JRT-PINC	0.899	0.899	0.895	0.887	0.874
	CJRT-PINC-θ	0.894	0.894	0.888	0.880	0.866
	CJRT-PINC-*m*	0.893	0.894	0.889	0.879	0.865
CJRT-PINC-θ	JRT-PINC	0.917	0.916	0.913	0.907	0.896
	CJRT-PINC-θ	0.956	0.955	0.954	0.953	0.949
	CJRT-PINC-*m*	0.938	0.945	0.952	0.950	0.937
CJRT-PINC-*m*	JRT-PINC	0.895	0.893	0.891	0.881	0.866
	CJRT-PINC-θ	0.901	0.900	0.897	0.888	0.873
	CJRT-PINC-*m*	0.903	0.903	0.901	0.892	0.877

S5　实证研究补充结果

表 S5.1　研究 3 实证数据 Q 矩阵

题目	K_1	K_2	K_3	K_4	K_5	K_6	K_7
CM015Q01	0	1	0	0	1	0	0
CM015Q02D	1	0	0	0	1	0	0
CM015Q03D	1	0	0	0	1	0	0
CM020Q01	0	0	1	0	0	0	1
CM020Q02	0	0	1	0	0	0	1
CM020Q03	0	0	1	0	0	0	1
CM020Q04	0	0	1	0	0	0	1
CM038Q03T	0	0	0	1	0	1	0
CM038Q05	0	0	0	1	0	1	0
CM038Q06	0	0	0	1	0	1	0

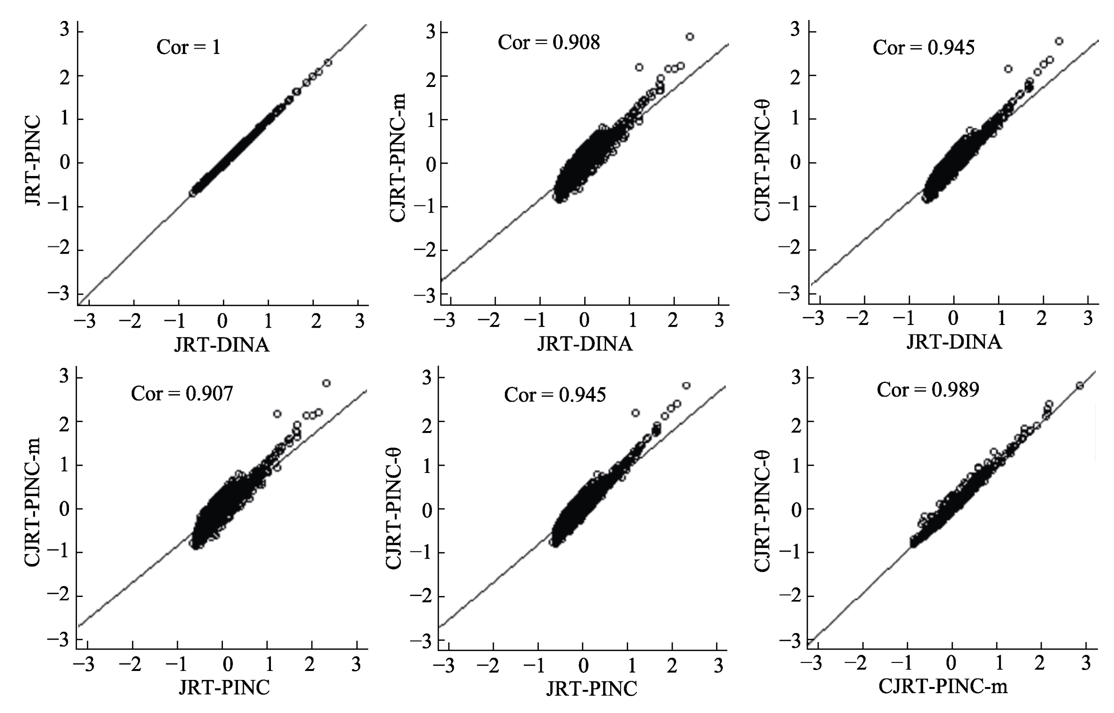


图 S5.1　实证数据中模型加工速度参数估计值散点图.
注: y 轴对应模型比 x 轴对应模式的估计值更大，则散点趋势高于对角线表明；反之，散点趋势低于对角线。

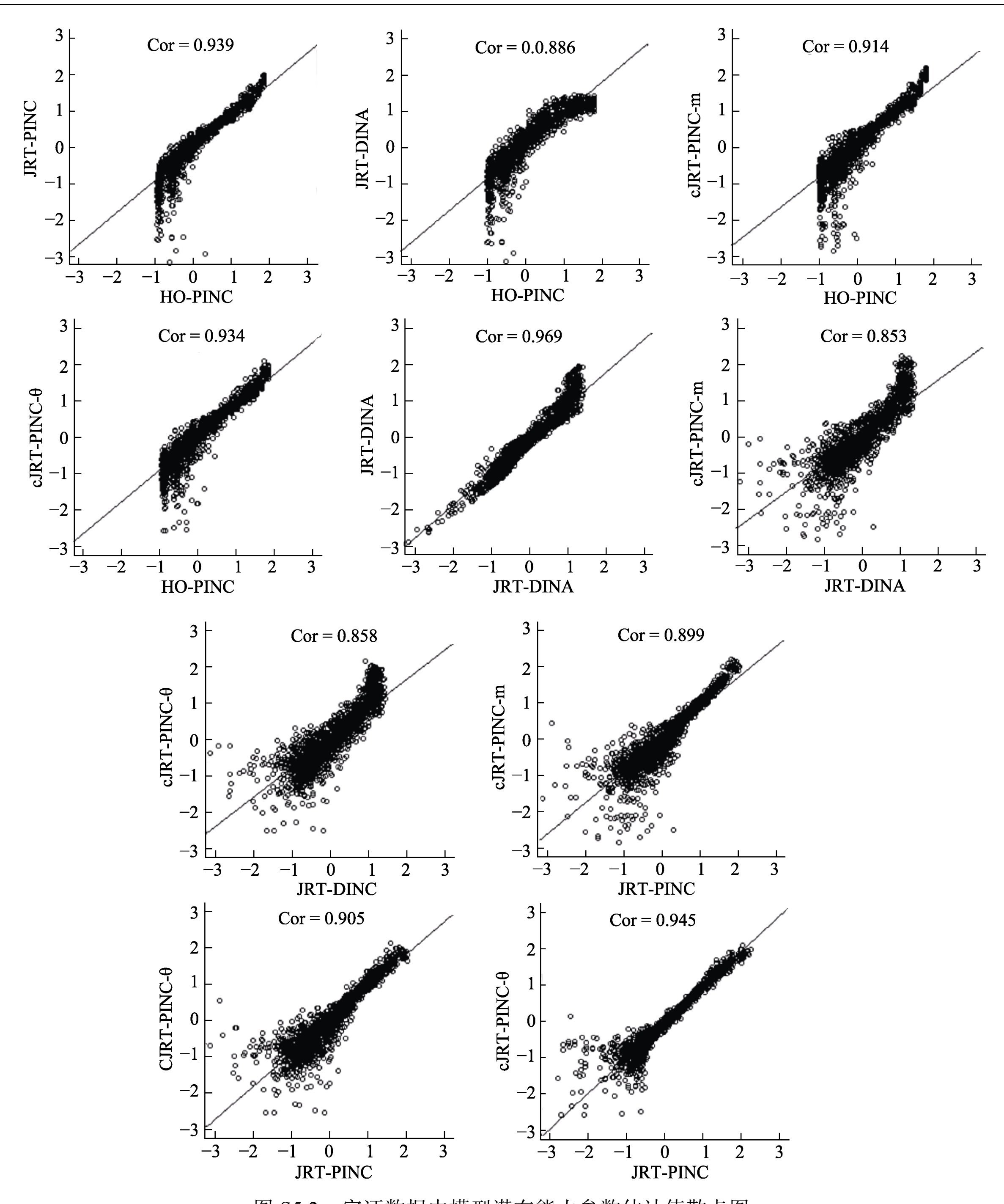


图 S5.2　实证数据中模型潜在能力参数估计值散点图.

注: y 轴对应模型比 x 轴对应模式的估计值更大, 则散点趋势高于对角线表明; 反之, 散点趋势低于对角线。